IMMUNOGLOBULIN BINDING PROTEIN ARRAYS IN EUKARYOTIC CELLS

CROSS-REFERENCE TO RELATED PATENT APPLICATIONS

This application is a continuation of US Application Serial No. 09/563,222, filed 05/02/2000, incorporated herein by reference in its entirety.

BACKGROUND OF THE INVENTION

Technical Field

The present invention relates generally to arrays of immunoglobulin binding proteins. The invention is more particularly related to methods for the expression of arrays of foreign immunoglobulin binding proteins in eukaryotic cells, such as plant cells, as well as to transformed eukaryotic cells that express such arrays.

Background of the Invention

Immunoglobulin molecules play key roles in a variety of physiological processes. Such molecules, which include antibodies and portions thereof, are critical for immune system function, and have found numerous therapeutic and diagnostic applications. The discovery of immunoglobulin molecules with desired binding characteristics is the focus of many current drug discovery efforts.

Traditional techniques for immunoglobulin molecule discovery involve the expression of a multitude of immunoglobulin molecule genes in an array of hybridoma cells, other forms of immortalized B-lymphocytes or phage-infected bacteria. For monoclonal antibody expression, individual antibody-producing B-lymphocytes from an immunized animal are generally fused with cells derived from an immortalized B-lymphocyte tumor. Clones of hybrid cells are then screened to identify those that grow indefinitely and secrete the desired immunoglobulin molecule. The polynucleotides encoding the monoclonal antibodies can then be isolated and used to express all or part of the antibody in other organisms, such as bacteria, yeast and plants. The ability to express immunoglobulins on the surface of bacteriophage has enabled the generation of immunoglobulin libraries that could represent all possible combinations of heavy and light chains derived from any population of B-lymphocytes. These libraries have been used successfully to identify high affinity combining sites recognizing a wide variety of antigens. A significant drawback to this technique is the

randomization of heavy and light chain combining sites requiring the generation of very large numbers of recombinant phage to identify specific heavy and light chain binding pairs. This combinatorial aspect of random libraries makes expression of these libraries in other organisms unfeasible. Newer technologies involve transgenic mice expressing antibodies from human chromosomal segments which can be used to generate hybridoma arrays expressing human antibodies.

Arrays formed in B-lymphocytes, phage infected bacteria or transgenic animals have been useful within certain immunoglobulin molecule screens, but difficulties have been encountered with producing large quantities of immunoglobulin molecules in these cells. Large-scale production of immunoglobulin molecules from any of the traditional organisms is typically very expensive. Further, phage infected bacteria are incapable of providing the variety of immunoglobulin molecule structures that may be desired. Similarly, the usefulness of transgenic animal cells has been limited by the susceptibility of such cells to infection with viruses or other microorganisms.

For economic and other reasons, it would be desirable to use genetically engineered plants as the primary vehicle for the discovery of immunoglobulin molecules, as well as for the ultimate production of immunoglobulin molecules to be used in industrial, clinical or research applications. The advantages of plants for production of immunoglobulin molecules include a low cost of production, relatively low capital investment compared to fermentation systems, the absence of animal viruses and prions, production of the immunoglobulin molecule in a biochemical background of defined proteins such as seed proteins, ease of storage and transport, and a facile scale-up to unlimited quantities of raw material. It would also be desirable to be able to express a library of binding proteins that is not derived from a combinatorial process of randomly paired heavy and light chains.

It is known that immunoglobulin molecules can be expressed in a variety of eukaryotic hosts including plant cells. A wide variety of structural genes have been isolated from mammalian cells and viruses, joined to transcriptional and translational initiation and termination regulatory signals from a source other than the structural

gene, and introduced into plant hosts in which these regulatory signals are functional. Among those host cells that have been transformed with individual immunoglobulin molecule-encoding nucleic acids are monocots (*e.g.*, corn, rice and wheat), dicots (*e.g.*, tobacco, soybean, alfalfa, petunia, and *Arabidopsis*) and lower plants (*e.g.*, *Chlamydomonas*). Plants transformed with nucleic acids encoding individual immunoglobulin molecules have been able to produce fully functional and fully assembled immunoglobulins (see Hiatt et al., *Nature 342*:76-78, 1989; Firek et al., *Plant Molecular Biology 23*:861-870, 1993; Van Engelen et al., *Plant Molecular Biology 26*:1701-1710, 1994; Ma et al., *Science 268*:716-719, 1995; Magnuson et al., *Protein Expression and Purification 7*:220-228, 1996; Schouten et al., *Plant Molecular Biology 30*:781-793, 1996; Fiedler et al., *Immunotechnology 3*:205-216, 1997; Verch et al., *J. Immunol. Meth. 220*:69-75, 1998; Zeitlin et al., *Nature Biotechnology 16*:1361-1364, 1998; DeJaeger et al., *Eur. J. Biochem. 259*:426-434, 1999; Fischer et al., *Biol. Chem. 380*:825-839, 1999; Khoudi et al., *Biotechnology and Bioengineering 64(2)*:135-143, 1999; McCormick et al., *Proc. Natl. Acad. Sci. USA 96*:703-708, 1999; Russell, *Curr. Top. Microbiol. Immunol. 240*:119-138, 1999).

In previous plant cell transformations, the transforming nucleic acid introduced a single immunoglobulin molecule. For example, tobacco plants have been transformed with individual gamma or kappa chains to produce individual plants expressing immunoglobulin molecule components. The respective tobacco transformants were then cross-pollinated to produce plants expressing a single antibody, wherein covalent bond formation between the two components resulted in the formation of enhanced binding capacity. In another instance, an antibody molecule was introduced into a single plant using a single vector. The vector encoded two immunoglobulin component chains and resulted in the formation, in the plant, of an immunoglobulin molecule comprising covalently linked heavy and light immunoglobulin chains.

Plant cells have not been used to express a diversity of immunoglobulin molecules in an array. As noted above, the ability to prepare an array of immunoglobulin molecules in plants or plant cells would facilitate identification of

useful immunoglobulin molecules and would enable a rapid transition from immunoglobulin molecule discovery to full scale production in a single organism.

Accordingly, there remains a need in the art for methods for generating arrays of immunoglobulin molecules in plants and plant cells, as well as other eukaryotic organisms and cells. The present invention fulfills these needs and further provides other related advantages.

Summary of the Invention

Briefly stated, the present invention provides methods for the production of arrays of biologically or physiologically active immunoglobulin binding proteins in eukaryotic cells. Within certain aspects, methods are provided for preparing an immunoglobulin binding protein array in plant cells, comprising the steps of: (a) transforming a population of plant cells with a library of at least two different polynucleotides encoding different immunoglobulin binding protein (IgBP) polypeptides that: (i) specifically bind to a ligand with a $K_D<10^{-6}$ moles/liter; or (ii) form one or more disulfide bonds with one or more polypeptides in the transfected cell, to generate a binding protein that specifically binds to a ligand with a $K_D<10^{-6}$ moles/liter; wherein the IgBP polypeptides (i) comprise four framework regions (*e.g.*, human or murine) alternating with three complementarity determining regions and (ii) comprise at least one peptide sequence having at least 75%, preferably at least 95%, sequence identity to a framework region of a native IgM, IgG, IgA, IgD, IgE, IgY, kappa or lambda immunoglobulin molecule; and wherein the IgBP polypeptides are not detectably expressed by the plant cells prior to transformation; and (b) selecting transformed plant cells, and therefrom preparing an IgBP array in plant cells. Each IgBP polypeptide may be a functional IgBP or an IgBP component (*e.g.*, a portion of an immunoglobulin molecule selected from the group consisting of heavy chains and fragments thereof, light chains and fragments thereof, J chains and secretory components) that, upon disulfide linkage to one or more IgBP components encoded by other polynucleotides in the library, forms a functional IgBP. Within certain specific

embodiments, a library employed in such methods commprises at least 10, 100, 1,000 or 10,000 different polynucleotides.

Within certain embodiments, such methods further comprise the step of: (c) growing the transformed plant cells on a growth medium that supports replication of the plant cells, such that functional IgBPs are assembled by the plant cells. Within other specific embodiments, such methods further comprise the steps of: (c) growing the transformed plant cells on a growth medium to form plants; and (d) sexually crossing the plants with themselves or other plants to generate progeny, such that the progeny comprise polynucleotides encoding IgBP components sufficient to form a functional IgBP. Such progeny may be seeds, or may be plants or plant cells that assemble functional IgBPs, and the IgBP polypeptides may, but need not, be secreted from the plant cells.

The present invention further provides, within other aspects, methods for preparing a heavy chain binding protein array in eukaryotic cells (*e.g.*, plant, insect or mammalian cells), comprising the steps of: (a) transforming a population of eukaryotic cells with a library of at least two different polynucleotides, wherein each polynucleotide encodes a different heavy chain binding protein (C_HBP) polypeptide that: (i) comprises an amino acid sequence that is at least 75% identical to a constant region tailpiece of a mu or alpha chain of a native immunoglobulin heavy chain; (ii) comprises multiple combining sites, wherein all of the combining sites satisfy the same one of the following requirements: (1) at least 75% identity to a 25 consecutive amino acid portion of an immunoglobulin light chain variable region or (2) at least 75% identity to a 25 consecutive amino acid portion of an immunoglobulin heavy chain variable region; and (iii) either (1) specifically binds to a ligand with a $K_D<10^{-6}$ moles/liter; or (2) forms one or more disulfide bonds with one or more polypeptides in the transfected cell, to generate a C_HBP that specifically binds to a ligand with a $K_D<10^{-6}$ moles/liter; and (b) growing the transformed cells on a medium that permits assembly of C_HBPs, wherein each C_HBP comprises at least four combining sites; and therefrom preparing a C_HBP array in eukaryotic cells. The polynucleotides may, for example, encode immunoglobulin alpha or mu chains. Within certain embodiments, the cells are

further transformed with one or more polynucleotides encoding polypeptides having sequences that are at least 75% identical to a sequence of an immunoglobulin J chain. Resulting C_HBPs may be assembled, for example, from four alpha chains and one J chain, from twelve mu chains and/or from ten mu chain and at least one J chain. C_HBPs or components thereof may, but need not, further comprise one or more portions of immunoglobulin molecules selected from the group consisting of J chains, secretory components and light chain constant regions. The C_HBPs may accumulate in an intracellular compartment of the cells or may be secreted from the cells.

Within further aspects, methods are provided for preparing a heavy chain binding protein array in eukaryotic cells, comprising the steps of: (a) exposing multiple copies of a polynucleotide encoding a native heavy chain to a mutagen, such that random or site-directed mutagenesis of the polynucleotide occurs, resulting in a library of heavy chain variants; (b) transforming a population of eukaryotic cells with the library of heavy chain variants; and (c) growing the transformed cells on a medium that permits assembly of C_HBPs, wherein each C_HBP comprises at least four combining sites; and therefrom preparing a C_HBP array in eukaryotic cells.

Methods are further provided for preparing a plant C_HBP array, comprising the steps of: (a) transforming a population of plant cells with a library of at least two different polynucleotides, wherein each polynucleotide encodes a different C_HBP component that forms one or more disulfide bonds with one or more polypeptides in the transformed cell to generate a C_HBP that specifically binds to a ligand with a $K_D<10^{-6}$ moles/liter, wherein each component: (i) comprises an amino acid sequence that is at least 75%, preferably at least 95%, identical to a constant region tailpiece of a mu or alpha chain of a native immunoglobulin heavy chain; and (ii) comprises multiple combining sites, wherein all of the combining sites satisfy the same one of the following requirements: (1) at least 75% identity to a 25 consecutive amino acid portion of an immunoglobulin light chain variable region or (2) at least 75% identity to a 25 consecutive amino acid portion of an immunoglobulin heavy chain variable region; (b) growing the transformed plant cells on a growth medium to form plants; and (c) sexually crossing the plants to generate progeny, such that the progeny

comprise polynucleotides encoding C_HBP components sufficient to form a functional C_HBP that comprises at least four combining sites; and therefrom preparing a plant C_HBP array. The progeny may be seeds, or may be plants or plant cells that assemble functional C_HBPs. Within certain specific embodiments, a library employed in such methods commprises at least 10, 100, 1,000 or 10,000 different polynucleotides. The C_HBPs may accumulate in an intracellular compartment of the cells or may be secreted from the cells.

Within further aspects, the present invention provides C_HBP arrays in eukaryotic cells, comprising at least two eukaryotic cells (*e.g.*, plant, insect or mammalian cells) that are each transformed with a different polynucleotide encoding at least one C_HBP polypeptide that: (a) comprises an amino acid sequence that is at least 75% identical to a constant region tailpiece of a mu or alpha chain of a native immunoglobulin heavy chain; (b) comprises multiple combining sites, wherein all of the combining sites satisfy the same one of the following requirements: (i) at least 75% identity to a 25 consecutive amino acid portion of an immunoglobulin light chain variable region or (i) at least 75% identity to a 25 consecutive amino acid portion of an immunoglobulin heavy chain variable region; (c) either (i) specifically binds to a ligand with a $K_D<10^{-6}$ moles/liter; or (ii) forms one or more covalent bonds with one or more polypeptides in the transfected cell, to generate a C_HBP that specifically binds to a ligand with a $K_D<10^{-6}$ moles/liter; and (d) differs in amino acid sequence from other C_HBPs in the array; wherein the cells assemble C_HBPs comprising at least four combining sites. The polynucleotides may, for example, encode polypeptide components of immunoglobulin molecules independently selected from the group consisting of heavy chains and fragments thereof, light chains and fragments thereof, J chains and secretory components. Within certain specific embodiments, the cells in the array assemble at least 10, 100, 1,000 or 10,000 different polynucleotides. Also within certain embodiments, each cell in such an array may be transfected with at least two different polynucleotides, each encoding a different C_HBP component, such that each cell assembles a functional C_HBP comprising the C_HBP components.

Within further aspects, the present invention provides compositions comprising an array of encapsulated C_HBPs, wherein each C_HBP: (a) comprises an amino acid sequence that is at least 75% identical to a constant region tailpiece of a mu or alpha chain of a native immunoglobulin heavy chain; (b) comprises at least four combining sites, wherein all of the combining sites satisfy the same one of the following requirements: (i) at least 75% identity to a 25 consecutive amino acid portion of an immunoglobulin light chain variable region; or (ii) at least 75% identity to a 25 consecutive amino acid portion of an immunoglobulin heavy chain variable region; and (c) either (i) specifically binds to a ligand with a $K_D<10^{-6}$ moles/liter; or (ii) forms one or more covalent bonds with one or more polypeptides in a cell, to generate a C_HBP that specifically binds to a ligand with a $K_D<10^{-6}$ moles/liter; and (d) differs in amino acid sequence from other C_HBPs in the array.

Methods are further provided for preparing a heavy chain binding protein array in eukaryotic cells, comprising the steps of: (a) exposing multiple copies of a polynucleotide encoding a native heavy chain to a mutagen, such that random or site-directed mutagenesis of the polynucleotide occurs, resulting in a library of heavy chain variants; (b) transforming a population of eukaryotic cells with the library of heavy chain variants; and (c) growing the transformed cells on a medium that permits assembly of C_HBPs, wherein each C_HBP comprises at least four combining sites; and therefrom preparing a C_HBP array in eukaryotic cells.

Within further aspects, the present invention provides C_HBPs that: (a) comprise an amino acid sequence that is at least 75% identical to a constant region tailpiece of a mu or alpha chain of a native immunoglobulin heavy chain; (b) comprise at least four combining sites, wherein all of the combining sites satisfy the same one of the following requirements: (i) at least 75% identity to a 25 consecutive amino acid portion of an immunoglobulin light chain variable region; or (ii) at least 75% identity to a 25 consecutive amino acid portion of an immunoglobulin heavy chain variable region; and (c) either (i) specifically bind to a ligand with a $K_D<10^{-6}$ moles/liter; or (ii) form one or more covalent bonds with one or more polypeptides in a cell, to generate a C_HBP that specifically binds to a ligand with a $K_D<10^{-6}$ moles/liter.

These and other aspects of the present invention will become apparent upon reference to the following detailed description and attached drawings. All references disclosed herein are hereby incorporated by reference in their entirety as if each was incorporated individually.

Brief Description of the Drawing

Figures 1A-1B is a table depicting representative framework region (FR) sequences for human and mouse heavy and light chains, as indicated.

Detailed Description of the Invention

As noted above, the present invention is directed to arrays of immunoglobulin binding protein (IgBP) polypeptides in eukaryotic organisms or cells, such as plants and plant cells. Each cell within such an array is transformed with at least one polynucleotide encoding one or more IgBPs or polypeptide components thereof (either integrated within the nuclear genome or resident in the cytoplasm). Within certain embodiments, each cell comprises polynucleotides encoding multiple polypeptide components of one or more IgBPs, such that each cell is capable of assembling at least one functional (*i.e.*, biologically active) IgBP. Transformed cells may further comprise biologically active IgBPs, although in some cases (*e.g.*, seeds generated from transformed plants), the functional IgBP may not be generated. Transgenic arrays embodying the present invention are useful for the discovery of novel IgBPs. Within certain embodiments, IgBPs within arrays provided herein are not derived from a combinatorial process of randomly paired heavy and light chains. It has been found, within the context of the present invention, that plants and other eukaryotic cells are capable of assembling IgBPs comprising any variable region (from light or heavy chain) expressed as specific heavy chain isotypes.

In certain embodiments, IgBP arrays may be generated in high yields in transformed, sexually reproducible plants. Transgenic plants provided herein are generally morphologically normal, but for the presence of one or more copies of the foreign polynucleotides in some or all of their cells. The respective gene products can be present in substantially all or a portion of the plant cells (*e.g.*, the products can be

localized to a cell type, tissue or organ). From the transgenic arrays, individual organisms or clones of cells can be identified rapidly, enabling the easy access to an economical, high yield process for the large scale production of a desired IgBP.

GLOSSARY

Prior to setting forth the present invention in detail, definitions of certain terms used herein are provided.

Immunoglobulin binding protein (IgBP): An immunoglobulin binding protein (i) comprises an amino acid sequence that is at least 75% identical to at least one framework region of a native immunoglobulin molecule (*e.g.*, IgM, IgG, IgA, IgD, IgE, IgY kappa or lambda) and (ii) is a functional binding protein. Framework regions are described below, under "Immunoglobulins." A protein P is a functional binding protein if (1) for one molecular, ionic or atomic ligand A the K_D (P, A)<10^{-6} moles/liter (preferably <10^{-7} moles/liter), where K_D (X,Y) = [X][Y]/[X:Y], and (2) for a different molecular, ionic or atomic species B, K_D (P, B)>10^{-4} moles/liter. Such a protein P is said to specifically bind A. Immunoglobulin binding proteins (IgBPs) generally function as a binding protein by virtue of the properties of a sequence of amino acids comprising a combining site, as defined below. An IgBP may comprise a single immunoglobulin chain or fragment thereof, multiple identical immunoglobulin chains or fragments thereof, or multiple non-identical immunoglobulin chains or fragments thereof. IgBPs include, for example, single chain antigen binding proteins, Fabs and Fvs. Also included are heavy chain binding proteins (C_HBPs), discussed in greater detail below.

Component of an IgBP: a polypeptide capable of forming one or more covalent bonds (preferably disulfide bonds) with one or more other polypeptides to generate a functional binding protein. A component is not itself a functional binding protein. For example, a multimeric antibody is considered an IgBP, and the polypeptide chains that are joined by covalent bonds to form an antigen binding site are considered to be IgBP components. Examples of such components include but are not

limited to heavy chains and fragments thereof, light chains and fragments thereof, J chain and fragments thereof, and secretory component and fragments thereof.

IgBP polypeptide: encompasses both functional IgBPs and IgBP components.

Immunoglobulin heavy chain binding protein (C_HBP): an IgBP that (i) comprises multiple combining sites derived from (*i.e.*, at least 75% identical to at least 25 consecutive amino acids of) either immunoglobulin light chain or heavy chain variable regions, but not both; and (ii) comprises a native heavy chain constant region sequence, or a fragment or other variant thereof, provided that the amino acid sequence of such a component is at least 75% identical to a constant region tailpiece (defined below) of a mu or alpha chain of a native immunoglobulin heavy chain. A C_HBP that comprises combining sites derived from one or more heavy chain variable regions does not comprise a combining site derived from a light chain variable region. Similarly, a C_HBP that comprises combining sites derived from one or more light chain variable regions does not comprise a combining site derived from a heavy chain variable region. Multiple C_HBP components may be covalently linked to generate a functional C_HBP, or a single polypeptide may be sufficient. Representative C_HBPs include proteins assembled from four alpha chains and one J chain, from twelve mu chains or from ten mu chains and at least on J chain.

Heavy chain binding protein (C_HBP) component: a polypeptide that is capable of forming one or more covalent bonds (preferably disulfide bonds) with one or more other polypeptides to generate a functional C_HBP. The component is not itself a binding protein. Examples of such components include, but are not limited to, heavy chains and fragments thereof, and J chain and fragments thereof.

Heavy chain binding protein (C_HBP) polypeptide: includes both C_HBPs and C_HBP components.

Immunoglobulin: any of the structurally related proteins or glycoproteins that function as antibodies. Polypeptides can be determined to comprise an immunoglobulin sequence based on sequence homology to known heavy chains, lights chains, J chains and related immunoglobulin sequences (*see, e.g.*, Kabat et al.,

Sequences of Immunological Interest, National Institutes of Health, Bethesda, Md. 1991). In general, an immunoglobulin sequence should be at least 95% identical to a known immunoglobulin sequence over at least 50 consecutive amino acid residues of a constant region, if the immunoglobulin contains a constant region, and 75% identical to at least 25 consecutive amino acids of a variable region. Immunoglobulins may comprise multiple immunoglobulin components. Examples of such components include, but are not limited to, heavy chains and fragments thereof, light chains and fragments thereof, J chain and fragments thereof, and secretory component and fragments thereof. An immunoglobulin is generally identified by its binding specificity for a unique epitope.

Immunoglobulins are composed of the linear combination of a basic domain structure. Each domain contains two beta-pleated sheets, one beta sheet consisting of four beta strands, the other consisting of three beta strands. The two beta sheets are covalently linked by a disulfide bond. Antibody variable regions contain three sequences termed complementarity determining regions (CDR) within which are amino acid sequences of high variability when comparing numerous variable region sequences. Flanking each CDR are sequences of lesser variability termed framework regions (FR), of which there are four. The positions of the CDRs primarily coincide with the loops between beta strands, and conversely the FRs correlate with the beta strands themselves of the basic domain structure. For example, CDR1 (closest to the amino terminus of the immunoglobulin polypeptide) lies between beta strands 4-2 and 3-1, and CDR2 is between 3-1 and 4-4. The three stranded beta sheets of variable regions are the contact areas between the light chain and heavy chain variable regions. The following table illustrates the typical structure of variable regions.

Table 1

Amino Acid Residues Associated with Framework Regions and Complementarity Determining Regions of Immunoglobulin Light and Heavy Chain Variable Domains

Segment	Light chain amino acids	Heavy chain amino acids
FR1	1-23	1-30

CDR1	24-34	31-35
FR2	35-49	36-49
CDR2	50-56	50-65
FR3	57-88	66-94
CDR3	89-97	95-102
FR4	98-107	103-113

A polypeptide that is "at least 75% identical to a framework region of a native IgM, IgG, IgA, IgD, IgE, IgY, kappa or lambda immunoglobulin molecule" generally comprises a sequence that retains at least 75% amino acid sequence identity to a native FR region. The significance of similarity may be determined statistically using a computer program such as the Align program described by Dayhoff et al., *Meth. Enzymol. 91*:524-545, 1983. Representative FR sequences are presented in Figures 1A-1B, and others may be found, for example, in Kabat et al., Sequences of Immunological Interest, National Institutes of Health, Bethesda, Md. 1991.

Table 2 shows representative consensus sequences for mouse and human CDRs. The specific sequences of individual CDRs can be found in Kabat et al., Sequences of Immunological Interest, National Institutes of Health, Bethesda, Md. 1991. General rules for identifying a CDR are as follows: (1) CDR1 (light chain): start is approximately residue 24; residue before is always a Cys; residue after is always a Trp (typically TRP-TYR-GLN, but also, TRP-LEU-GLN, TRP-PHE-GLN, TRP-TYR-LEU); length 10 to 17 residues; (2) CDR 2 (light chain): start is always 16 residues after the end of CDR 1; residues before generally ILE-TYR, but also, VAL-TYR, ILE-LYS, ILE-PHE; length always 7 residues (except 7FAB which has a deletion in this region); (3) CDR 3 (light chain): start is always 33 residues after end of CDR 2 (except 7FAB which has the deletion at the end of CDR-L2); residue before is always Cys; residues after always PHE-GLY-XXX-GLY (SEQ ID NO: 1) (4) CDR 1 (heavy chain): start is approx residue 26 (always 4 after a CYS); residues before always CYS-XXX-XXX-XXX (SEQ ID NO: 2); residue after always a TRP; typically TRP-VAL, but also, TRP-ILE, TRP-ALA; (5) CDR 2 (heavy chain): start is always 15 residues

after the end of CDR 1; residues before typically LEU-GLU-TRP-ILE-GLY (SEQ ID NO: 3), but there are a number of variations; residues after LYS/ARG-LEU/ILE/VAL/PHE/THR/ALA-THR/SER/ILE/ALA; (6) CDR 3 (heavy chain): start is always 33 residues after end of CDR 2 (always 2 after a CYS); residues before always CYS-XXX-XXX (typically CYS-ALA-ARG); residues after always TRP-GLY-XXX-GLY (SEQ ID NO: 4).

Table 2

CDR consensus sequences

Light chain sequences.

Type	CDR 1	CDR 2	CDR 3
Human kappa I	RASQSLVSISSYLA (SEQ ID NO: 5)	AASSLES (SEQ ID NO: 6)	QQYNSLPEWT (SEQ ID NO: 7)
Human kappa II	RSSQSLLHSDGDTYLN (SEQ ID NO: 8)	LVSNRAS (SEQ ID NO: 9)	MQALQPRT (SEQ ID NO: 10)
Human kappa III	RASQSVSSSYLA (SEQ ID NO: 11)	GASSRAT (SEQ ID NO: 12)	QQYGSSPPLT (SEQ ID NO: 13)
Human kappa IV	KSSQSVLYSSNNKNYLA (SEQ ID NO: 14)	WASTRES (SEQ ID NO: 15)	QQYYSTPT (SEQ ID NO: 16)
Human lambda I	SGSSSNIIGNNYVS (SEQ ID NO: 17)	DNNKRPS (SEQ ID NO: 18)	ATWDDSLSANSAPV (SEQ ID NO: 19)
Human lambda II	TGTSSDVGGYNAVS (SEQ ID NO: 20)	DVTDRPS (SEQ ID NO: 21)	SSYGGGSNV (SEQ ID NO:22)
Human lambda III	SGDNLGDKYVH (SEQ ID NO: 23)	DDNKRPS (SEQ ID NO:24)	QAWDSSSDHPGVV (SEQ ID NO:25)
Mouse kappa I	KSSQSLLNSGNQKNYLA (SEQ ID NO: 26)	WASTRES (SEQ ID NO:27)	QNDYSYPLT (SEQ ID NO: 28)
Mouse kappa II	RSSQSLVHSNGNTYLE (SEQ ID NO: 29)	KVSNRFS (SEQ ID NO:30)	FQGTHVPPYT (SEQ ID NO: 31)
Mouse kappa III	RASESVDSYGNSFMH (SEQ ID NO: 32)	AASNLES (SEQ ID NO: 33)	QQSNEDPPWT (SEQ ID NO: 34)
Mouse kappa IV	SASSSVSSSYLH (SEQ ID NO: 35)	RTSNLAS (SEQ ID NO: 36)	QQWSSYPGLT (SEQ ID NO: 37)
Mouse kappa V	RASQDDISNYLN	YASRLHS	QQGNTLPPRT

	(SEQ ID NO: 38)	(SEQ ID NO: 39)	(SEQ ID NO: 40)
Mouse kappa IV	SASSSVSYMH (SEQ ID NO: 41)	DTSKLAS (SEQ ID NO: 42)	QQWSSNPPMPLT (SEQ ID NO: 43)

Heavy chain sequences.

Type	CDR 1	CDR 2	CDR 3
Human heavy I	SYAIS (SEQ ID NO: 44)	WINPYGNGDTNYAQKFQG (SEQ ID NO: 45)	APGYGSGGGCYRGDYFDY (SEQ ID NO: 46)
Human heavy II	SYGWSWN (SEQ ID NO: 47)	RIYYRAYSGSTTYNPSLKS (SEQ ID NO: 48)	ELPGGYTGDDYYYGSGFDV (SEQ ID NO: 49)
Human heavy III	SYAMS (SEQ ID NO: 50)	VISGKTDGGSTYYADSVKG (SEQ ID NO: 51)	GRPGDSLSGYYYYYHYFDY (SEQ ID NO: 52)
Mouse heavy I	SGYWNNS (SEQ ID NO: 53)	YISGYSGSTYYNPSLKS (SEQ ID NO: 54)	GGYGYGYYYYDYYYYFDY (SEQ ID NO: 55)
Mouse heavy II	DYYMNN (SEQ ID NO: 56)	DINPGNGGTSYNQKFKG (SEQ ID NO: 57)	GSYYSSSYMAYYAFDY (SEQ ID NO: 58)
Mouse heavy III	DFYME (SEQ ID NO: 59)	ASRNKANDYTTEYSASVKG (SEQ ID NO: 60)	DYYYGSSYYEGPVYWYFDV (SEQ ID NO: 61)

Immunoglobulin superfamily molecule: a molecule that has a domain size and amino acid residue sequence that is significantly similar to immunoglobulin or immunoglobulin related domains. The significance of similarity may be determined statistically as discussed above. A sequence identity of greater than 75% with the immunoglobulins listed in Kabat et al. (Sequences of Immunological Interest, National Institutes of Health, Bethesda, Md. 1991) indicates a member of the immunoglobulin superfamily. Also, a typical Align score of less than 3 indicates that the molecule being tested is a member of the immunoglobulin gene superfamily.

Combining site: a portion of a binding protein made up of those amino acid residues that contact the ligand or antigen by ionic interactions, hydrogen bonding, Van der Waals interaction or hydrophobic interaction. A combining site in a typical binding protein is located in a variable region of light and/or heavy chains and, more specifically, in a CDR of the light and/or heavy chains. A combining site need not comprise more than one polypeptide. A combining site generally comprises four FR

and three CDR sequences, which alternate as displayed in Table 1. The precise amino acids present within a combining site may generally be determined by x-ray diffraction analysis of the binding protein with bound ligand or antigen. See, for example, Amit et al., *Science 233*:4765, 747-53, 1986.

Immunoglobulin constant region: a portion of an immunoglobulin polypeptide that follows the carboxy terminus of the variable region. This is usually in the vicinity of amino acid #108 in light chains and amino acid #114 in heavy chains. Constant regions determine the isotype designation of the immunoglobulin and include but are not limited to kappa or lambda light chain constant regions and gamma, mu, alpha, epsilon and delta heavy chain constant regions. Constant regions of heavy chains are divided into domains. The first domain following the variable region is designated CH1. Domains following CH1 include the hinge region, CH2, CH3 and possibly CH4 and membrane spanning domain. Examples of immunoglobulin constant regions can be found in Kabat et al., Sequences of Immunological Interest, National Institutes of Health, Bethesda, Md. 1991.

Tailpiece of a constant region: tailpiece regions are located after the CH3 or CH4 segments of IgA or IgM constant regions, respectively. Examples of tailpieces are listed in Table 3. Other tailpiece regions may be identified based on similarity in sequences to one of the representative heavy chain tailpiece regions in Table 3. In general, a tailpiece should be at least 50% identical to a sequence in Table 3 and always contains a cysteine as the penultimate carboxy terminus amino acid. When co-expressed in appropriate eukaryotic cells containing an endomembrane system, tailpiece regions attached to either alpha or mu constant regions are capable of forming a disulfide bond between the penultimate cysteine of the tailpiece and a cysteine in J chain. Formation of these disulfide bonds can result in the polymerization of alpha or mu constant regions.

Table 3

Representative Heavy Chain Tailpiece Regions

Isotype	Species	Sequence

IgA	Human	PTHVNVSVVMAEVDGTCY (SEQ ID NO: 62)
IgA	Mouse	PTVNVIMSEGDGICY (SEQ ID NO: 63)
IgA	Mouse	PTNVSVSVIMSEGDGICY (SEQ ID NO: 64)
IgA	Rat	PTNVNVSVIMSEGDGICY (SEQ ID NO: 65)
IgA	Rabbit	PTHVNVSVVVADVEAVCY (SEQ ID NO: 66)
IgM	Human	PTLYNVSLVMSDTAGTCY (SEQ ID NO: 67)
IgM	Human	PTLYNVSLIMSDTGGTCY (SEQ ID NO: 68)
IgM	Hamster	PTLYNVSLIMSDAGGTCY (SEQ ID NO: 69)
IgM	Hamster	PTLYNVSLVLSDTAGZCY (SEQ ID NO: 70)
IgM	Rat	PTLYNVSLIMSDTASTCY (SEQ ID NO: 71)
IgM	Chicken	PSFVNVSLVLMDTVNSCN (SEQ ID NO: 72)

J chain: a polypeptide that is substantially identical (*i.e.*, at least 80% identical) in sequence to a J chain of Kabat et al. (Sequences of Immunological Interest, National Institutes of Health, Bethesda, Md. 1991). J chains are further capable of forming disulfide bonds to the penultimate cysteine residues in the tailpiece region of alpha or mu constant regions, thereby forming a polymeric structure, such as a dimeric IgA or a pentameric IgM. Disulfide bonds between J chain and alpha or mu heavy chain constant regions are generally formed in the endomembrane system of a cell

during the process of secretion. The presence of disulfide bonds can be measured by comparison of the polypeptides in the presence or absence of an appropriate reducing agent such as dithiothreitol or mercaptoethanol. Comparative analysis of peptides can be accomplished by, for example, denaturing gel electrophoresis using SDS and polyacrylamide as described in Molecular Cloning: A Laboratory Manual, Sambrook et al., eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1989) and Antibodies: A Laboratory Manual, Harlow and Lane, eds., Cold Spring Harbor, N.Y. (1988).

Alpha chain: a polypeptide that is substantially identical (*i.e.*, at least 90% identical) in sequence to the constant region of an alpha chain of Kabat et al. (Sequences of Immunological Interest, National Institutes of Health, Bethesda, Md. 1991).

Mu chain: a polypeptide that is substantially identical (*i.e.*, at least 90% identical) in sequence to the constant region of a mu chain of Kabat et al. (Sequences of Immunological Interest, National Institutes of Health, Bethesda, Md. 1991).

Secretory component: a polypeptide that binds to polymeric immunoglobulins containing J chains. Secretory components are derived from the polyimmunoglobulin receptor. The sequences of some secretory components and polyimmunoglobulin receptors have been determined. Other polypeptide that share at least 75% sequence identity with a known secretory component, and that retain the ability to bind to polymeric immunoglobulins containing J chains, are also considered to be secretory components.

Heavy chain: a polypeptide that comprises an amino acid sequence that is at least 90% identical to the constant region of a native heavy chain sequence and an amino acid sequence that is at least 75% identical to the variable region of a native heavy chain sequence (*see* Kabat et al., Sequences of Immunological Interest, National Institutes of Health, Bethesda, Md. 1991) and that, when co-expressed with a light chain in the endomembrane system of an appropriate eukaryotic cell is capable of forming a heavy chain-light chain complex, potentially an antibody, joined by disulfide bonding. Heavy chains may be identified as the larger of the two polypeptides present

within a divalent antibody, Fab, Fab'2 or Fv. Heavy chains are also found as components of polyvalent antibodies such as IgAs and IgMs.

Light chain: a polypeptide that is at least 90% identical to the constant region of a native light chain sequence and 75% identical to the variable region of a native light chain sequence (*see* Kabat et al., Sequences of Immunological Interest, National Institutes of Health, Bethesda, Md. 1991) and that when co-expressed with a heavy chain in the endomembrane system of an appropriate eukaryotic cell is capable of forming a heavy chain-light chain complex, potentially an antibody, joined by disulfide bonding. Light chains may be identified as the smaller of the two polypeptides present within a divalent antibody, Fab, Fab'2 or Fv. Light chains are also found as components of polyvalent antibodies such as IgAs and IgMs.

Light chain or heavy chain variable region: a region of a light chain or heavy chains that contains three complementarity determining regions (CDR), flanked by four framework regions (FR), as discussed above. A variable region generally has two antiparallel β-pleated sheets, one of which consists of four β strands, and one of which consists of three β strands. In an antibody the three-stranded sheet is the contact area between heavy chain and light chain variable regions. The two β-pleated sheets are covalently linked by a disulfide bond (the half cystines that form this bond are conserved). The CDRs coincide with the loops between β strands, while the FRs correlate with the β strands themselves. An IgBP that does not comprise a light chain variable region generally does not contain a polypeptide that contains CDRs or FRs but may contain the constant region of a light chain.

IgBP array: a population of eukaryotic cells or organisms (*e.g.*, plants) that are transformed with different polynucleotides, each of which encodes a different IgBP or polypeptide component thereof. Each cell or organism is transformed with at least one such polynucleotide, and preferably at least two such polynucleotides, such that the array comprises at least two organisms or cells that are transformed with different polynucleotides. In general, the IgBP polypeptides encoded by the polynucleotides should not be detectably expressed by untransformed cells or organisms. An array may be displayed on a growth surface that allows for the

replication of the cells or organisms. An array may be present, for example, in plants or plant cells, or may be stored in seed form.

Dicotyledon (dicot): A flowering plant whose embryos have two seed halves or cotyledons. Examples of dicots are: tobacco; tomato; the legumes including alfalfa; oaks; maples; roses; mints; squashes; daisies; walnuts; cacti; violets; and buttercups.

DNA: Deoxyribonucleic acid.

Epitope: A portion of a molecule that is specifically recognized by an immunoglobulin. It is also referred to as the determinant or antigenic determinant.

Eukaryotic hybrid vector: A DNA molecule by means of which DNA coding for a polypeptide (insert) can be introduced into a eukaryotic cell.

Fab fragment: A polypeptide consisting of a portion of an antibody molecule containing the active portions of an antibody heavy chain and an antibody light chain covalently coupled together and capable of specifically combining with antigen. Fab fragments are typically prepared by proteolytic digestion of substantially intact antibody molecules with papain using methods that are well known in the art. However an Fab fragment may also be prepared by expressing in a suitable host cell the desired portions of an antibody heavy chain and light chain using methods well known in the art.

Fv fragment: A polypeptide consisting of the active portions of an antibody heavy chain variable region and an antibody light chain variable region covalently coupled together and capable of specifically combining with antigen. Fv fragments are typically prepared by expressing in a suitable host cell the desired portions of antibody heavy chain variable region and light chain variable region using methods well known in the art.

Mutagenesis: A process whereby the nucleotide sequence of an original polynucleotide is changed in one or a few locations to produce derivative polynucleotides of substantially the same sequence. Mutagenesis can be accomplished by manipulation of polynucleotides *in vitro* by, for example, using various commonly available enzymes and mutagenic oligonucleotides. Mutagenesis can also be

accomplished *in vivo* using the immune system of an animal to introduce desired changes in polynucleotides encoding immunoglobulins for example. B cell maturation, for example, is a process whereby a mutated polynucleotide is derived from an original polynucleotide.

Insert: A DNA sequence foreign to the host, consisting of a structural gene and optionally additional DNA sequences.

(Selective) Genetic marker: A DNA sequence coding for a phenotypical trait by means of which transformed cells can be selected from untransformed cells.

Leader sequence: A contiguous series of amino acids preceding a polypeptide. The leader sequence may be cleaved from the polypeptide during the process of secretion.

Signal Sequence: A DNA sequence coding for an amino acid sequence attached to the polypeptide which binds the polypeptide to the endoplasmic reticulum and is essential for protein secretion.

Lower plant: Any non-flowering plant including ferns, gymnosperms, conifers, horsetails, club mosses, liverworts, hornworts, mosses, red algae, brown algae, gametophytes, sporophytes of pteridophytes, and green algae (*e.g., Chlamydomonas*).

Monocotyledon (monocot): A flowering plant whose embryos have one cotyledon or seed leaf. Examples of monocots are: lilies; grasses; corn; rice; grains, including oats, wheat and barley; orchids; irises; onions and palms.

Open reading frame: A sequence of nucleic acids when read sequentially three at a time (triplets) contains no stop codon sequences. Stop codon sequences are UAG, UAA, and UGA.

Paratope: An antigen binding site of an antibody molecule.

Plant cell: A cell that depends for its growth on light and contains chloroplasts.

Pollination: the transfer of pollen from male to female flower parts. "Self-pollination" refers to the transfer of pollen from male flower parts to female flower parts on the same plant. This process typically produces seed from which progeny plants can be grown. "Cross-pollination" is the transfer of pollen from the

male flower parts of one plant to the female flower parts of another plant. This process typically produces seed from which viable progeny can be grown.

Polypeptide and peptide: A series of amino acid residues covalently connected one to the other by peptide bonds between the alpha-amino and carboxy groups of adjacent residues or by disulfide bridges between two cysteines.

Protein: A linear series of greater than about 50 amino acid residues connected one to the other as in a polypeptide.

Promoter: A recognition site on a DNA sequence or group of DNA sequences that provide an expression control element for a gene and to which RNA polymerase specifically binds and initiates RNA synthesis (transcription) of that gene. An "inducible promoter" is a promoter where the rate of RNA polymerase binding and initiation is modulated by external stimuli. Such stimuli include light, heat, anaerobic stress, alteration in nutrient conditions, presence or absence of a metabolite, presence of a ligand, microbial attack, wounding and the like. A "viral promoter" is a promoter with a DNA sequence substantially similar to the promoter found at the 5' end of a viral gene. A typical viral promoter is found at the 5' end of the gene coding for the p21 protein of MMTV described by Huang et al., *Cell 27*:245, 1981. A "synthetic promoter" is a promoter that was chemically synthesized rather than biologically derived. Usually synthetic promoters incorporate sequence changes that optimize the efficiency of RNA polymerase initiation. A "constitutive promoter" is a promoter where the rate of RNA polymerase binding and initiation is approximately constant and relatively independent of external stimuli. Examples of constitutive promoters include the cauliflower mosaic virus 35S and 19S promoters described by Poszkowski et al., *EMBO J. 3*:2719, 1989 and Odell et al., *Nature 313*:810, 1985. A "temporally regulated promoter" is a promoter where the rate of RNA polymerase binding and initiation is modulated at a specific time during development. Examples of temporally regulated promoters are given in Chua et al., *Science 244*:174-181, 1989. A "spatially regulated promoter" is a promoter where the rate of RNA polymerase binding and initiation is modulated in a specific structure of the organism such as the leaf, stem or root. Examples of spatially regulated promoters are given in Chua et al., *Science*

244:174-181, 1989. A "spatiotemporally regulated promoter" is a promoter where the rate of RNA polymerase binding and initiation is modulated in a specific structure of the organism at a specific time during development. A typical spatiotemporally regulated promoter is the EPSP synthase-35S promoter described by Chua et al., *Science 244*:174-181, 1989.

RNA: Ribonucleic acid.

rRNA: Ribosomal RNA.

Secretion signal: A contiguous series of amino acids attached to or contained within a polypeptide which enables secretion of the polypeptide from a cell. The amino acids may or may not be cleaved from the polypeptide during the process of secretion.

Seed: A viable dormant embryo enclosed in a testa that is derived from the integument(s). The embryo is associated with food reserve organs.

Single-chain antigen-binding (SCAB) protein: A polypeptide composed of an immunoglobulin light-chain variable region amino acid sequence (V_L) tethered to an immunoglobulin heavy-chain variable region amino acid sequence (V_H) by a peptide that links the carboxyl terminus of the V_L sequence to the amino terminus of the V_H sequence or by a peptide that links the amino terminus of the V_L sequence to the carboxyl terminus of the V_H sequence.

Single-chain antigen-binding (SCAB) protein-coding gene: A recombinant gene coding for a single-chain antigen-binding protein.

Structural gene: A gene coding for a polypeptide and being equipped with a suitable promoter, termination sequence and optionally other regulatory DNA sequences, and having a correct reading frame.

T-DNA: A segment of transferred DNA specific for plants.

Ti-plasmid: Tumor-inducing plasmid specific for plants.

Ti-DNA: A segment of DNA from Ti-plasmid.

Transfection: A process whereby foreign DNA is introduced into a eukaryotic cell. The process may or may not result in cellular transformation.

Transformation: A process whereby foreign DNA is introduced into the genome or cytoplasm of an organism resulting in the expression of a trait that did not previously exist in the organism. Examples of added traits are resistance to toxic chemicals and expression of foreign proteins not normally produced by the organism.

IMMUNOGLOBULIN BINDING PROTEINS (IGBP) POLYPEPTIDES

As noted above, IgBPs comprise one or more polypeptides in which amino acids are linked by covalent peptide bonds. In general, an IgBP (i) comprises an amino acid sequence that is at least 75% identical to at least one framework region of a native immunoglobulin molecule (*e.g.*, IgM, IgG, IgA, IgD, IgE, IgY, kappa or lambda) and (ii) is a functional binding protein. Sequence identity may be determined using any of a variety of well known algorithms, which may be readily optimized by those of ordinary skill in the art. One such algorithm is employed by the Align program described by Dayhoff et al., *Meth. Enzymol. 91*:524-545, 1983. A functional binding protein, as discussed above, specifically binds to a ligand with a $K_D<10^{-6}$ moles/liter (preferably $<10^{-7}$ moles/liter). K_D may be readily determined using well known assays.

An IgBP may comprise a single immunoglobulin chain or fragment thereof, multiple identical immunoglobulin chains or fragments thereof, or multiple non-identical immunoglobulin chains or fragments thereof. IgBPs include, for example, single chain antigen binding proteins, Fabs and Fvs. Other IgBPs are heavy chain binding proteins (C_HBPs), which comprise multiple combining sites composed of amino acid residues derived from the constant region of an immunoglobulin heavy chain and a variable region from any source (*e.g.*, either heavy or light chain, but not both). In a preferred embodiment the variable region is derived from a heavy chain. C_HBPs further comprise an amino acid sequence that is at least 75% identical to a constant region tailpiece of a mu or alpha chain of a native immunoglobulin heavy chain. A C_HBP may further comprise one or more J chains, which can serve to link other component polypeptides. Representative C_HBPs include proteins assembled from four alpha chains and one J chain, from twelve mu chains or from ten mu chains and at least on J chain.

IgBPs may be made up of component polypeptides linked by covalent bonds, preferably disulfide bonds. Preferred components comprise one or more portions of immunoglobulin molecules such as heavy chains and fragments thereof, light chains and fragments thereof, J chains and fragments thereof and secretory components and fragments thereof. A component polypeptide may comprise a native immunoglobulin sequence or a variant of such a sequence. As noted above, an "IgBP polypeptide" is any polypeptide that is a functional IgBP or an IgBP component.

For preparation of an array, a library of polynucleotides encoding IgBP polypeptides is generally employed. Such a library generally comprises at least two different polynucleotides (preferably at least 10, 100, 1,000 or 10,000 different polynucleotides), each of which encodes a different IgBP or component thereof. An IgBP is different from another IgBP if there are one or more differences between the amino acid sequences of the IgBPs. Preferably, such differences result in differences in the affinity of the IgBP for a respective substrate, ligand or epitope. Preferred libraries encode variants of an original polynucleotide. Such a library may be derived from the original polynucleotide by a process of mutagenesis that may occur *in vitro* (*e.g.*, using commonly available enzymes and mutagenic oligonucleotides) or *in vivo* (*e.g.*, by B cell maturation). Such a library generally comprises polynucleotides encoding variants that differ from an original immunoglobulin by one or more amino acid substitutions and/or deletions, such that each variant retains at least 75% identity, preferably at least 95% identity, to the original immunoglobulin. Certain such libraries may contain polynucleotides encoding immunoglobulin variants that differ from the original immunoglobulin only in one or more point mutations. Polynucleotides in a library are preferably present within a vector that facilitates transformation of target eukaryotic cells, and subsequent expression of binding protein or component thereof in transformed cells.

There are a variety of sources for nucleic acids encoding populations of immunoglobulin binding proteins. For example, immunoglobulins may be derived from the B cells of an immunized host. Such a population of immunoglobulin binding proteins is derived from a multitude of cells each expressing a different

immunoglobulin binding protein. Populations of immunoglobulin binding proteins or potential immunoglobulin binding proteins can also be derived from the mutagenesis of unique sequences of immunoglobulins.

Methods for isolating polynucleotides encoding a population of IgBPs are well known in the art. See, for example, Molecular Cloning: A Laboratory Manual, Sambrook et al., eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1989) and Antibodies: A Laboratory Manual, Harlow and Lane, eds., Cold Spring Harbor, N.Y. (1988). As noted above, particularly preferred polynucleotides encode immunoglobulin heavy and light chain variable regions, or portions thereof. Such polynucleotides may be isolated from cells obtained from a vertebrate, preferably a mammal, which has been immunized with an antigenic ligand (antigen) against which activity is sought (*i.e.*, a preselected antigen). The immunization can be carried out conventionally and antibody titer in the animal can be monitored to determine the stage of immunization that corresponds to the affinity or avidity desired. Partially immunized animals typically receive only one immunization and cells are collected therefrom shortly after a response is detected. Fully immunized animals display a peak titer that is achieved with one or more repeated injections of the antigen into the host mammal, normally at two to three week intervals. Usually three to five days after the last challenge, the spleen is removed and the genes coding for immunoglobulin heavy and immunoglobulin light chains are isolated from the rearranged B cells present in the spleen using standard procedures. See Current Protocols in Molecular Biology, Ausubel et al., eds., John Wiley and Sons, New York (1987) and Antibodies: A Laboratory Manual, Harlow and Lane, eds., Cold Spring Harbor, N.Y. (1988).

In addition to the spleen, rearranged B cells can be derived from the bone marrow of individuals who have been exposed to specific antigens. For example, patients in hospitals can be exposed to infectious organisms that they would not ordinarily encounter. Frequently, these patients mount an immune response that results in B cell maturation and deposition in bone marrow of B cells expressing antibodies that neutralize the pathogen. In addition to nosocomial exposure, memory B cells are deposited in response to a variety of infections, including HIV, HPV, HSV and CMV.

Genes en coding V_H and V_L polypeptides can be derived from cells producing IgA, IgD, IgE, IgG or IgM, and most preferably from IgG producing cells. Methods for preparing fragments of genomic DNA from which immunoglobulin variable region genes can be cloned are well known in the art. See for example, Herrmann et al., *Methods in Enzymol. 152*:180-183, 1987; Frischauf, *Methods in Enzymol. 152*:183-190, 1987; and Frischauf, *Methods in Enzymol. 152*:199-212, 1987.

Probes useful for isolating polynucleotides encoding immunoglobulin products include the sequences encoding constant portions of the V_H and V_L, sequences encoding the framework regions of V_H and V_L and probes for the constant region of the entire rearranged immunoglobulin gene, these sequences being obtainable from available sources. See, for example, Early and Hood, Genetic Engineering, Setlow and Hollaender eds., Vol. 3:157-188, Plenum Publishing Corporation, New York (1981); and Kabat et al., Sequences of Immunological Interest, National Institutes of Health, Bethesda, Md. (1991). Polynucleotides encoding a polypeptide subunit or component of an IgBP can be isolated from either the genomic DNA containing the gene expressing the polypeptide or the messenger RNA (mRNA) that codes for the polypeptide. The use of mRNA is preferred, due to the difficulty in juxtaposing sequences of genomic DNA that encode the polypeptide, where the sequences are separated by introns. In such cases, the DNA fragment(s) containing the proper exons must be isolated, the introns excised, and the exons spliced together in the proper order and orientation. Methods for isolating mRNA coding for peptides or proteins are well known in the art. See, for example, Current Protocols in Molecular Biology, Ausubel et al., John Wiley and Sons, New York (1987); "Guide to Molecular Cloning Techniques", in Methods In Enzymology, Volume 152, Berger and Kimmel, eds. (1987); and Molecular Cloning: A Laboratory Manual, Sambrook et al., eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1989).

Mutagenesis can be used to generate a multitude of polynucleotides encoding different variants of native IgBPs, which can then be resolved by expression in an array of eukaryotic cells or plant cells or plants as provided herein. Mutant proteins may be obtained in which one or more specific changes in the codons of a gene

of choice are introduced. One or a few polypeptides are then expressed for evaluation of binding properties. At the other extreme is random mutagenesis, by means of relatively nonspecific changes of codons at a variety of sites in the gene of choice. Arrays of plants and plant cells, as well as other eukaryotic cells, can be used for functional screens of mutant IgBPs. Preferably, polynucleotides within such arrays encode variants that differ from a native IgBP sequence in one or more amino acid substitutions and/or deletions, such that each variant retains at least 75%, preferably at least 95%, identity to the native IgBP. Particularly preferred are arrays in which each polynucleotide differs from the native IgBP in one or more point mutations. Mutagenesis is also a process that occurs naturally during the development of antibodies. During the course of B cell maturation, antibody-encoding genes are recombined, selected, and mutated to produce plasma cells that encode antibodies with higher affinity for an antigen compared to the original polynucleotide encoding the antibody with specificity for the same antigen.

Regardless of the precise type of IgBP, array polynucleotides are formulated so as to permit entry into, and replication within, a target host cell. For certain transfection techniques, the polynucleotides are cloned into a suitable expression vector. Any vector can be used for such transfection, provided that the vector is capable of the transcription of IgBP genes as well as selectable markers in a target cell. Typical expression vectors useful for expression of genes in plants are well known in the art and include vectors derived from the tumor-inducing (Ti) plasmid of *Agrobacterium tumefasciens* described by Rogers et al., *Meth. in Enzymol. 153*:253-277, 1987. However, several other expression vector systems are known to function in plants. See, for example, Register et al., *Plant Mol Biol. 25*:951-961, 1994; Verch et al., *J. Immunol. Meth. 220*:69-75, 1998.

Alternatively, for certain transfection techniques an expression vector is unnecessary or undesirable. In these cases, either single or multiple DNA fragments containing the polypeptide coding genes linked to plant expression control elements (expression cassettes) are introduced directly into target plant cells. The DNA

fragment used to transfect plant cells may include either single or multiple polypeptide coding genes.

Suitable expression vectors and cassettes generally contain expression control elements, which include the promoter. The polypeptide coding genes are operatively linked to the expression vector or cassette to allow the promoter sequence to direct RNA polymerase binding and synthesis of the desired polypeptide coding gene. Useful in expressing the polypeptide coding gene are promoters that are inducible, viral, synthetic, constitutive, temporally regulated, spatially regulated and/or spatiotemporally regulated. The choice of expression vector, and ultimately the selection of a promoter to which the polypeptide coding gene is operatively linked, depends directly (as is well known in the art) on the functional properties desired (*e.g.*, the location and timing of protein expression and the host cell to be transformed). However, an expression vector or cassette useful in practicing the present invention is at least capable of directing the replication, and preferably also the expression, of the polypeptide coding gene included in the DNA segment to which it is operatively linked.

In preferred embodiments, the expression vector or cassette used to express the polypeptide-coding genes includes a selection marker that is effective in a target cell, preferably a drug resistance selection marker. Alternatively, a separate vector or expression cassette comprising a selection marker may be co-transfected. A preferred drug resistance marker is the gene whose expression results in kanamycin resistance (*e.g.*, the chimeric gene containing the nopaline synthase promoter, Tn5 neomycin phosphotransferase II and nopaline synthase 3' nontranslated region described by Rogers et al., in Methods For Plant Molecular Biology, a Weissbach and H. Weissbach, eds., Academic Press Inc., San Diego, Calif. (1988)). Useful plant expression vectors include pMON 530, pKYLX and pUC. Another preferred drug resistance marker is the gene for phosphinothricin acetyltransferase, PAT, which confers resistance to Liberty herbicide. Alternatively, the selection marker could permit the growth of the transfected cells or tissue on a alternative substrate present in the medium (*e.g.*, xylA gene; Haldrup et al., *Plant Mol Bio 37(2)*:287-296, 1998).

A variety of methods have been developed to operatively link DNAs to vectors via complementary cohesive termini. For instance, complementary homopolymer tracks can be added to the DNA segment to be inserted and to the vector DNA. The vector and DNA segment are then joined by hydrogen bonding between the complementary homopolymeric tails to form recombinant DNA molecules. Alternatively, synthetic linkers containing one or more restriction endonuclease sites can be used to join the DNA segment to the expression vector. The synthetic linkers are attached to blunt-ended DNA segments by incubating the blunt-ended DNA segments with a large excess of synthetic linker molecules in the presence of an enzyme that is able to catalyze the ligation of blunt-ended DNA molecules, such as bacteria phage T4 DNA ligase. Thus, the products of the reaction are DNA segments carrying synthetic linker sequences at their ends. These DNA segments are then cleaved with the appropriate restriction endonuclease and ligated into an expression vector that has been cleaved with an enzyme that produces termini compatible with those of the synthetic linker. Synthetic linkers containing a variety of restriction endonuclease sites are commercially available from a number of sources including New England BioLabs, Beverly, MA. Polymerase chain reaction (PCR) can also be used to introduce appropriate restriction sites into polynucleotide sequences as would be needed for ligating the polynucleotide into a restriction site in a plant or eukaryotic cell expression vector.

IMMUNOGLOBULIN BINDING PROTEINS (IGBP) ARRAYS

A library of polynucleotides as described above may be used to generate an array in eukaryotic cells or organisms (*e.g.*, plants or seeds) using standard transfection techniques appropriate for the cell or organism of interest. In general, a library is used to transfect a population of eukaryotic cells such that some or all of the cells contain one or more polynucleotides encoding IgBP polypeptides that are not be detectably expressed by untransfected cells. Such transfection can result in functional transformation, which permits the cells to produce IgBPs and/or components thereof. The cells may then be grown on an appropriate medium to allow for replication and the

functional expression of IgBPs or components thereof. Optionally, a progeny population may be created from a sexual cross of transformants expressing IgBP components, such that some or all of the progeny population contain polynucleotides encoding sufficient IgBP components to form functional IgBPs. Also optionally, progeny populations may be created from sexual crosses of transformants expressing multiple IgBPs such that each member of the progeny population expresses a single functional IgBP. Sexual crosses between plants expressing IgBPs, or between wild-type plants and plants expressing IgBPs, results in the genetic segregation of the various IgBPs among the various progeny seeds or progeny plants. As such, some of the progeny will express a single IgBP.

Transformed organisms are generally morphologically normal but for the presence of the foreign genes in some or all of their cells. The genes can be present in one copy or multiple copies in some or all of the cells. The respective gene products can be present in substantially all or a portion of the cells (*i.e.*, the products can be localized to a cell type, tissue or organ). Transformed cells may be dormant (*e.g.*, non-germinating seeds), present in culture or in an intact organism.

Transgenic cells or organisms (such as plants) may be produced, in one embodiment, by introducing into the nuclear genome a mammalian gene or genes that code for a multiplicity of IgBPs. In another embodiment, the introduced gene or genes remain extrachromasomal after being introduced by viral infection. In each single organism or population of cells, one or more copies of the nucleic acid encoding the IgBP polypeptide(s) is integrated into the genome or is resident in the cytoplasm. Since a multiplicity of open reading frames is used in the transformation, one or more nucleic acids encoding IgBP polypeptides with different specificities can be introduced into each cell or population of cells. Transformation techniques that result in multiple copies of nucleic acids integrated into the nucleus of a plant cell include *Agrobacterium* mediated transformation, biolistic transformation, electroporation, solid particle intrusion, lipofection, chemically-induced DNA uptake, microinjection or macroinjection. Viral infection can result in a multiplicity of nucleic acids functionally operative in the cell cytoplasm.

In general, recombinant immunoglobulins can be prepared by isolating DNA fragments corresponding to the heavy and light chain variable regions of a monoclonal antibody and joining them to each other by any one of the standard methods known to those of skill in the art and described by Molecular Cloning: A Laboratory Manual, Sambrook et al., eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1989). In one preferred embodiment, only DNA fragments corresponding to heavy chains are used to prepare suitable vectors for cell transformation. In one example of cell transformation, the recombinant DNA fragments can be inserted into *Agrobacterium* transfer vectors such that the genes of interest are inserted into the *Agrobacterium* genome. The recombinant *Agrobacterium* is then used to infect plant cells resulting in the production of the polypeptide of interest. In another example of cell transformation, the recombinant DNA fragments can be inserted into baculovirus transfer vectors such that the genes of interest are inserted into the viral genome in lieu of the baculovirus polyhedron gene. The recombinant virus is then used to infect insect cells resulting in the production of the polypeptide of interest.

An advantage of using insect cells that utilize recombinant baculoviruses for the production of IgBPs is that the baculovirus system allows production of mutant antibodies as well as combinatorial expression of immunoglobulin with other polypeptides (such as J chain) much more rapidly than stably transfected mammalian or plant cell lines. In addition, insect cells have been shown to correctly process and glycosylate eukaryotic proteins. Finally, the baculovirus expression of foreign protein has been shown to constitute as much as 50-75% of the total cellular protein late in viral infection, making this system an excellent means of producing milligram quantities of recombinant immunoglobulins.

The use of baculovirus *Autographica californica* nuclear polyhedron virus (AcNPV) and recombinant viral stocks in *Spodoptera frugiperda* (Sf9) cells to prepare large quantities of protein has been described by Smith et al., *Proc. Natl. Acad. Sci. U.S.A.* 82: 8404-8408, 1985, and Summers and Smith, *Bulletin B – Texas Agricultural Experiment Station*, May, 1987. A preferred method of preparing recombinant heavy chains covalently linked to J chain or the light chain constant region

is through the expression of DNA encoding recombinant heavy chain, J chain, and light chain constant region via the baculoviral expression system in Sf9 insect cells.

Isolated DNA fragments that encode preferred genes are then inserted into the baculovirus transfer vectors. A preferred transfer vector is based on pAc360. The DNA fragments are digested, purified and ligated into unique restriction sites in the vector using standard techniques known to those of skill in the art and described by Molecular Cloning: A Laboratory Manual, Sambrook et al., eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1989).

Recombinant plasmid vectors are then co-transfected with linearized wild-type AcNPV into Sf9 cells. Cotransfection is preferentially accomplished using cationic liposomes, which are commercially available (*e.g.*, from Invitrogen). Sf9 cells infected with occlusion-negative, recombinant viruses are then identified and can be grown in the desired quantity and under appropriate conditions such that large quantities of the desired binding protein are produced.

Methods for introducing polypeptide coding genes into mammalian cells have been described and are generally familiar to one skilled in the art. Such methods include, but are not limited to, liposome-mediated transformation, calcium phosphate transfection, electroporation, diethylamoniethyl dextran carrier and viral infection. It will be readily apparent that an optimal method of introducing genes into a particular eukaryotic cell species may not necessarily be the most effective for another species.

The advantage of using certain viruses for mammalian cell transformation is that a variety of mammalian cell lines are potential targets of viral infection. Recombinant Sindbis virus for example, allows rapid, high-level expression of heterologous proteins in mammalian cell lines as well as avian, reptilian, mosquito and *Drosophila* cells. The virus inhibits host protein synthesis allowing rapid identification and purification of the transgenic protein.

Isolated DNA fragments are introduced into a vector containing an appropriate polylinker site downstream from a promoter and upstream from a poly-A tail. A useful vector is pSinRep5 (Invitrogen). The recombinant vector is then linearized using appropriate restriction sites downstream from the poly-A tail. RNA

transcripts are then produced (using, for example, the SP6 promoter) containing non-structural genes for *in vivo* replication of the recombinant RNA and the promoter and transgene of interest. The RNA is then used to transfect mammalian or other cells, such as baby hamster kidney (BHK) cells. Protein expression can be assayed 4-72 hours post infection. Alternatively, viral particles can be harvested and used for infection of another cell line.

In certain embodiments, an array comprises plants or cells transformed with polynucleotides encoding immunoglobulin heavy chains and immunoglobulin light chains, such that antibody molecules are produced. Within such arrays, individual immunoglobulin heavy and light chains produced by each plant or plant cell may associate with each other and assume a conformation having an antigen binding site specific for a preselected or predetermined antigen, as evidenced by its ability to be competitively inhibited. When the binding protein is an antigen binding protein, its affinity or avidity is generally greater than 10^5 M^{-1}, preferably greater than 10^6 M^{-1}, and more preferably greater than 10^8 M^{-1}. Immunoglobulins for use in such embodiments may generally be derived from the B cells of an immunized host, each of which express a different IgBP.

Similarly, an IgBP array may comprise polynucleotides that encode portions of immunoglobulin heavy chains and portions of immunoglobulin light chains. The individual immunoglobulin heavy and light chain portions in each plant or plant cell may associate with each other and assume a conformation having an antigen binding site specific for a preselected or predetermined antigen. The antigen binding site on a Fab fragment has a binding affinity or avidity similar to the antigen binding site on an immunoglobulin molecule. Likewise, the antigen binding site on a SCAB protein has a binding affinity or avidity similar to the antigen binding site on an immunoglobulin molecule. Alternatively, an IgBP array may comprise polynucleotides that encode immunoglobulin heavy and light chain fragments that may associate within the plant cell to form Fv fragments with a biologically active conformation that has a binding site specific for a preselected or predetermined antigen. The antigen binding

site on individual Fv fragments has an affinity or avidity for an antigen similar to the affinity displayed by the antigen binding site present on an immunoglobulin molecule.

Within certain preferred embodiments, IgBP arrays (preferably C_HBP arrays) may be prepared in plants, plant cells and/or seeds. Such arrays may be prepared in any plant or plant cell that can be conveniently transformed. Suitable plant types include, but are not limited to, *Arabidopsis,* tobacco, Lemna (duckweed), corn, rice and *Chlamydomonas*. Techniques for transforming these and other plant types are well known in the art, and include *Agrobacterium*-mediated plant transformation, protoplast transformation, gene transfer into pollen, injection into reproductive organs, gene transfer using a plant cell infective virus or bacterium, mechanical disruption of plant cells containing no cell walls in the presence of DNA and solid particulates (*e.g.,* glass beads), injection into immature embryos, acceleration into the plant cells on solid particles or fibers, or electroporation. As noted above, libraries of polynucleotides encoding more than one IgBP are used for transfection. Transfection may result in introduction of one or more polynucleotides into the nuclear genome of a plant cell. Alternatively, the introduced polynucleotide(s) may remain extrachromosomal after being introduced by viral infection. Preferably, the transfection technique results in a population of cells containing multiple functionally operative polynucleotides encoding IgBPs, or components thereof. Such techniques (that result in integration into the nuclear genome) include *Agrobacterium*-mediated transformation, biolistic transformation, direct DNA transfer, fiber-mediated transformation, microinjection, macroinjection and electroporation. Viral infection can result in multiple polynucleotides encoding IgBPs, or components thereof, functionally operative in the plant cell cytoplasm. It will be apparent that the selection of an optimal method for introducing genes will depend, in part, on the particular plant species.

The use of *Agrobacterium*-mediated expression vectors to introduce DNA into plant cells is well known in the art. See, for example, the methods described by Fraley et al., *Biotechnology 3*: 629, 1985 and Rogers et al., *Methods in Enzymology, 153*:253-277, 1987. Further, the integration of the Ti-DNA is a relatively precise process resulting in few rearrangements. The region of DNA to be transferred is

defined by the border sequences and intervening DNA is usually inserted into the plant genome as described by Spielmann et al., *Mol. Gen. Genet. 205*:34, 1986 and Jorgensen et al., *Mol. Gen. Genet. 207*:471, 1987. Modern *Agrobacterium* transformation vectors are capable of replication in *Escherichia coli* as well as *Agrobacterium*, allowing for convenient manipulations, as described by Klee et al., in *Plant DNA Infectious Agents*, T. Hohn and J. Schell, eds., Springer-Verlag, New York (1985) pp. 179-203. Further recent technological advances in vectors for *Agrobacterium*-mediated gene transfer have improved the arrangement of genes and restriction sites in the vectors to facilitate construction of vectors capable of expressing various polypeptide coding genes. The vectors described by Rogers et al., *Methods in Enzymology 153*:253, 1987, for example, have convenient multi-linker regions flanked by a promoter and a polyadenylation site for direct expression of inserted polypeptide coding genes, and are suitable for present purposes.

In those plant species where *Agrobacterium*-mediated transformation is efficient, this is the method of choice because of the facile and defined nature of the gene transfer (*see* Horsch et al., *Science 227*:1129-1231, 1985; Feldmann and Marks, *Mol. Gen. Genet. 208*:1-9, 1987; Chang et al., *Plant J.* 5:551-558, 1994; Bechtold et al., *Acad. Sci. Paris Science de la Vie 316*:1194-1199, 1993; Hansen and Chilton, *Curr Top Microbiol Immunol 240*:21-57, 1999). Few monocots appear to be natural hosts for *Agrobacterium*. However, transgenic plants may be produced in a variety of monocots using *Agrobacterium* vectors (as described, for example, by Bytebier et al., *Proc. Natl. Acad. Sci. U.S.A. 84*:5345, 1987; Hiei et al., *Plant J. 6(2)*:271-282, 1994; Ishida et al., *Nat. Biotechnol. 14(6)*:745-750, 1996; Hansen and Chilton, *Curr Top Microbiol Immunol 240*:21-57, 1999). Alternatively, vector-free or direct DNA transfer methods have been developed to transform a variety of plant species. As an example, transformation of plant protoplasts can be achieved using methods based on calcium phosphate precipitation, polyethylene glycol treatment, electroporation, microinjection, and combinations of these treatments. See, for example, Potrykus et al., *Mol. Gen. Genet. 199*:183, 1985; Lorz et al., *Mol. Gen. Genet. 199*: 178, 1985; Fromm et al., *Nature 319*:791, 1986; Uchimiya et al., *Mol. Gen. Genet. 204*:204, 1986; Callis et al.,

Genes and Development 1:1183, 1987; Marcotte et al., *Nature 335*:454, 1988; Roest and Gilissen, *Acta Bot Neerl 38*:1-23, 1989; Davey et al., *Plant Mol Biol 13*:273-285, 1989; Shimamoto et al., *Nature 338*:274-276, 1989; and Datta et al., *Bio/Technology 8*:736-740, 1990. Silicon carbide-mediated transformation may be used to generate stably transformed, fertile maize plants (Frame, et al., *The Plant Journal for Cell and Molecular Biology 6*:941-948, 1994) as well as tobacco cell cultures (Kaeppler, et al., *Theoretical and Applied Genetics 84*:560-566,1992), *Agrostis alba* cell cultures (Asano et al., *Plant Cell Reports 13*:243-246, 1994), and *Chlamydomonas reinhardtii* (Dunahay, *BioTechniques 15*:452-460, 1993). Application of these systems to different plant species often depends upon the ability to regenerate that particular plant species from protoplasts or callus. Illustrative methods for the regeneration of cereals from protoplasts are described, for example, in Fujimura et al., *Plant Tissue Culture Letters 2*:74, 1985; Toriyama et al., *Theor Appl. Genet. 73*:16, 1986; Yamada et al., *Plant Cell Rep. 4*:85, 1986; Abdullah et al., *Biotechnology 4*:1087, 1986. Register et al., *Plant Mol. Bio. 25(6)*:951-961, 1994; Blackhall N.W. et al., "Callus Initiation, Maintenance, and Shoot Induction in Rice," *in* Plant Cell Culture Protocols, Hall R.D. (editor), Humana Press (Totowa, NJ), 1999, pgs. 19-30.

To transform plant species that cannot be successfully regenerated from protoplast or callus, other ways to introduce DNA into intact cells or tissues can be used. For example, regeneration of cereals from immature embryos or explants can be effected as described by Dasil, *Biotechnology 6*:397, 1988. In addition, "particle gun" or high-velocity microprojectile technology can be used. Using such technology, DNA is carried through the cell wall and into the cytoplasm on the surface of small (0.525 μm) metal particles that have been accelerated to speeds of one to several hundred meters per second (as described in Klein et al., *Nature 327*:70, 1987; Klein et al., *Proc. Natl. Acad. Sci. U.S.A. 85*:8502, 1988; and McCabe et al., *Biotechnology 6*:923, 1988). The metal particles penetrate through several layers of cells and thus allow the transformation of cells within tissue explants. Metal particles may be used to successfully transform corn cells and to produce fertile, stably transformed plants in a variety of species including model plants such as tobacco (Klein et al., 1988), as well as

important crop species such as soybean (McCabe et al., 1988), maize (Fromm et al., *Bio/technology* 8: 833-839,1990; Gordon-Kamm et al, *The Plant Cell 2*: 603-618 1990), rice (Christou et al., *Bio/technology 9*: 957-962, 1991), barley (Wan and Lemaux, *Plant Physiology 104*: 37-48, 1994) and wheat (Vasil et al., *Bio/technology 10*: 662-674, 1992; Weeks et al., *Plant Physiology 102*: 1077-1084, 1993). Transformation of tissue explants eliminates the need for passage through a protoplast stage and thus speeds the production of transgenic plants.

DNA can also be introduced into plants by direct DNA transfer into pollen, as described by Zhou et al., *Methods in Enzymology 101*:433, 1983; Hess, *Intern Rev. Cytol. 107*:367, 1987; Luo et al., *Plant Mol. Biol. Reporter 6*:165, 1988; and Saunders et al., *Molecular Biotechnology 3*:181-190, 1995. Expression of polypeptide coding genes can be obtained by injection of the DNA into reproductive organs of a plant as described by Pena et al., *Nature 325*:274, 1987. DNA can also be injected directly into the cells of immature embryos, followed by the rehydration of desiccated embryos as described by Neuhaus et al., *Theor. Appl. Genet. 75*:30, 1987; and Benbrook et al., in *Proceedings Bio. Expo. 1986*, Butterworth, Stoneham, Mass., pp. 27-54 (1986).

Regardless of the method of transformation, transformed plant cells are generally grown on a suitable medium. Cells may be grown on medium that does not promote differentiation (*e.g.*, as calli), or may be grown such that plants are regenerated. For example, some lower plant species, such as *Chlamydomonas*, do not require regeneration and can express immunoglobulin binding proteins immediately after selection for the expression of a selectable marker, as a result of growth in an appropriate medium. Suitable media for a wide variety of plant cell types are well known to those of ordinary skill in the art. In general, transformed cells are selected based on expression of the selection marker, such that only those cells that contain at least one transfected polynucleotide encoding an immunoglobulin binding protein are permitted to grow.

Non-combinatorial arrays expressing heavy chain binding proteins may be generated using, for example, the MaxBac Baculovirus Expression System

(Invitrogen). Briefly, insect cells are co-transfected with recombinant plasmid or plasmids and linearized Bac-and-Blue DNA. A viral supernatant fluid is harvested after the appropriate infection time and the virus is used to display recombinant plaques on agar plates using techniques recommended by the manufacturer. Individual plaques can then by used for the evaluation of binding protein expression. Alternatively, heavy chain binding protein arrays may be generated using the Sindbis Expression System (Invitrogen), allowing the expression of polypeptides in baby hamster kidney (BHK) cells and the production of viral particles that can be used to infect a variety of other eukaryotic cell lines.

The regeneration of plants from either single plant protoplasts or various explants is well known in the art. See, for example, Methods for Plant Molecular Biology, A. Weissbach and H. Weissbach, eds., Academic Press, Inc. (San Diego, CA; 1988). This regeneration and growth process includes the steps of selection of transformant cells and shoots, rooting the transformant shoots and growth of the plantlets in soil.

The regeneration of plants containing the foreign gene introduced by *Agrobacterium tumefasciens* from leaf explants can be achieved as described by Horsch et al., *Science 227*:1229-1231, 1985. In this procedure, transformants are grown in the presence of a selection agent and in a medium that induces the regeneration of shoots in the plant species being transformed, as described by Fraley et al., *Proc. Natl. Acad. Sci. U.S.A. 80*:4803, 1983. This procedure typically produces shoots within two to four weeks, and these transformant shoots are then transferred to an appropriate root-inducing medium containing the selective agent and an antibiotic to prevent bacterial growth. Transformant shoots that rooted in the presence of the selective agent to form plantlets are then transplanted to soil to allow the production of roots. These procedures will vary depending upon the particular plant species employed, such variations being well known in the art.

An array may consist of a set of seeds generated from transformed plants. While such seeds have the potential to germinate into plants that express IgBPs, expression does not generally occur when the seeds are dormant (although some

expression may occur during seed development). Nonetheless, such a set of seeds is considered an array.

Within arrays, the IgBPs or components thereof may be secreted from cells or may accumulate in one or more intracellular compartments. Secreted proteins may or may not be contained by a cell wall. For example, the CW15 mutant of *Chlamydomonas* does not express a cell wall and IgBPs containing leader sequences are secreted directly into the growth medium. In this instance, the desired IgBP can be identified by assaying a portion of the growth medium for the desired characteristic. In higher plants, secreted IgBPs are generally contained by the cell wall and may accumulate in the apoplastic water. A desired IgBP can be assayed after disruption of the cell wall using any of a variety of standard mechanical techniques (*e.g.*, mortar and pestle homogenization).

Within certain embodiments, an array of IgBPs in plant cells is derived from a progeny population of plant cells (*i.e.*, plant cells resulting from the sexual cross of transformants expressing IgBP components). In such an array, some or all of the progeny population contain polynucleotides encoding IgBP components, such that functional IgBPs are expressed in the array. In such arrays, useful IgBP component polynucleotides include those that encode a second polypeptide that can autogenously associate with a first polypeptide in such a way as to form a biologically functional IgBP. Examples of such IgBP components are heavy and light chains of antibodies.

The sexual cross of different members of a plant species has been well described by Mendel in 1865 (an English translation of Mendel's original paper together with comments and a bibliography of Mendel by others can be found in Experiments In Plant Hybridization, Edinburgh, Scotland, Oliver Boyd, eds., 1965). When the plants are flowering plants, the sexual cross involves contacting viable pollen from one population of plants with the stigma of another population of plants of a sexually compatible species. When the plant cells are photosynthetic unicellular organisms (*e.g.*, *Chlamydomonas*), the sexual cross involves fusion of cells of + and – mating types, followed by meiosis and tetrad formation. Progeny from sexual crosses can comprise seeds or tetrads.

Progeny resulting from a sexual cross of transformed plant cells may, within certain embodiments, be resolved into an array in which some or all of the progeny contain polynucleotides encoding a single IgBP. Briefly, transformed plant cells expressing polynucleotides encoding multiple IgBPs can be sexually crossed to produce a progeny population. For higher plants, the sexual cross involves pollination as described above. The source of pollen can be either the same plant (self pollination) or a different plant (cross pollination). In some cases pollination is not required for seed development. The resulting progeny population contains the genetic segregation events of the polynucleotides encoding various IgBPs or components. Some of the progeny may contain polynucleotides encoding a single IgBP.

For haploid organisms such as *Chlamydomonas*, resolution of multiple binding events may also be accomplished by molecular genetic techniques. Nucleic acids encoding IgBPs can be identified by Southern blotting and isolated by PCR using known primers. The isolated nucleic acids can then be used for re-transformation of the haploid plant cell to produce a secondary array representing the resolved nucleic acids encoding IgBPs. In this process, vector DNA encoding IgBPs is amplified using oligonucleotides complementary to the boundary regions of the expression cassette. The amplified DNA is ideally suitable for ligation into an expression vector for re-transformations of a new population of *Chlamydomonas* cells. Transformants from this population contain relatively few different IgBPs and therefore there is a higher probability of generating a population of transformed cells wherein each nucleic acid is expressed uniquely by an individual transformed cell.

IMMUNOGLOBULIN BINDING PROTEIN ARRAY COMPOSITIONS

The present invention also provides compositions that comprise an array of encapsulated IgBPs. IgBPs may be encapsulated, for example, within plant cells, plant cell walls, enteric coatings, coatings and the like. Particularly preferred are compositions containing no more than 10,000 grams of plant material for each 100 nanograms of IgBP. Such compositions typically contain at least 100 nanograms of plant material for each 10 grams of IgBP. In more preferred embodiments, a

composition comprises no more than 10,000 grams of plant material for each milligram of IgBP, but at least 100 nanograms of plant material for each gram of IgBP and more preferably at least one milligram of plant material for each 500 milligrams of IgBP.

A composition may further comprise substances such as chlorophyll, synergistic compounds, medicines, compounds derived from medicinal plants and/or various pharmaceuticals.

ASSAYS FOR SCREENING ARRAYS OF IMMUNOGLOBULIN BINDING PROTEINS

Transgenic arrays as provided herein are useful for the discovery of IgBPs having desired characteristics. From the transgenic arrays provided herein, individual eukaryotic organisms or clones of eukaryotic cells can be identified rapidly, enabling the easy access to an economical, high yield process for the large scale production of a desired IgBP.

As noted above, plants are a particularly preferred host organism for the transgenic arrays described herein. When a sexual cross is performed between two populations of plant cells, each expressing IgBPs or components of IgBPs, some or all of the progeny can express functional IgBPs that comprise multiple IgBP components. Such progeny may be of particular use within functional screens for IgBPs having multiple polypeptide subunits.

More specifically, once regenerated plants are adapted to growth in soil, small sections of leaf can be removed to assay for the presence of IgBPs. Alternatively, the plants can be allowed to set seed by either self-pollination or cross pollination and the assays for expression of IgBPs can be performed on individual seeds or portions of seeds. When the transformants are individual cells (*e.g.*, *Chlamydomonas*) the cells may be grown in solid or liquid medium to obtain a sufficient quantity for performing assays. Assays can be immunological (*e.g.*, ELISA) or functional. Functional assays may involve, for example, binding of the IgBP to cognate factors immobilized on a solid support. When the IgBP is an antibody, the cognate factor is an antigen. Assays for determining the number of copies of IgBP as well as the number of different IgBPs being expressed in the primary transformants include Southern blotting and DNA

sequencing of cloned cDNAs derived from cellular RNA. Description of these assays may be found in Antibodies: A Laboratory Manual, Harlow and Lane, eds., Cold Spring Harbor, N.Y. (1988) and Molecular Cloning: A Laboratory Manual, Sambrook et al., eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1989).

The following Examples are offered by way of illustration and not by way of limitation.

EXAMPLES

Example 1

Isolation and Expression in Maize of Immunoglobulin Heavy Chain-Coding Genes And Immunoglobulin Light Chain-Coding Genes Derived From B Cells in Bone Marrow

This Example illustrates the preparation of representative immunoglobulin binding protein components for use in preparing an array of immunoglobulin binding proteins in plant cells.

Bone marrow (~10^7 cells) was obtained from donors at local hospitals. Donors were chosen on the basis of (a) having had *Clostridium difficile*-associated diarrhea within the previous two months and (b) having toxin neutralizing antibodies in their serum. Bone marrow was obtained from the subset of this population whose diarrhea resolved without the intervention of any further antibiotic therapy.

Bone marrow cells were washed multiple times in cold PBS then incubated with combined FITC-labeled *Clostridium difficile* toxin A and toxin B. This procedure was used to identify B cells expressing anti-toxin A or anti-toxin B on their cell surfaces. The *Clostridium difficile* toxins were prepared from cultures of *Clostridium difficile* as described by Lyerly et al., *Infect Immun 35*:1147-50, 1982.

Fluorescent labeling of the toxins was by reaction with fluorescein isothiocyanate (FITC) using procedures provided by the manufacturer (Pierce). Incubation of bone marrow cells with the combined FITC-labeled *Clostridium difficile* toxin A and toxin B was by the procedures described in Hoven et al., *J. Immunol. Methods 117*:2275-84, 1989. The tagged cells were shown to be B cells by double staining with phycoerythrin-conjugated anti-human Ig.

Total RNA from the selected bone marrow cells was prepared using the QIAamp RNA Blood Mini-Kit system (Qiagen) according to the instructions provided.

PCR amplification of the isolated RNA employed oligonucleotides complementary to conserved 3' and 5' regions of immunoglobulin (IgG1) transcripts as previously described in Antibody Engineering, Carl A, Borrebaeck, Ed., W.H. Freeman and Company, New York, N.Y., 1995. The oligonucleotide primers introduced unique

restriction sites at each end of the amplified polynucleotide to allow for ligation into a plant expression vector. The amplified portion of the immunoglobulin polynucleotides extended from the 5' end encoding the respective FR1 regions to the 3' end encoding the stop codon. In addition, appropriate 5' and 3' restriction sites were introduced in the process of amplification. The amplified heavy and light polynucleotide chains therefore corresponded to the first six codons at the 5' end of the RNAs and the last six codons of the constant regions with the restriction sites introduced outside of the respective terminal codons. This immunoglobulin binding protein array consisting of double stranded amplified cDNAs from the sorted cells was prepared for insertion into a plant expression vector, by converting the ends of the double stranded cDNA to the appropriate sticky ends for ligation into the vector (Molecular Cloning: A Laboratory Manual, Sambrook et al., eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. 1989) by digestion with a restriction endonuclease according to the manufacturer's protocol (New England BioLabs).

The vector allowed for the expression of both heavy and light chains using the ubiquitin promoter and nopaline synthase terminator as described by Ishida et al., *Nature Biotechnology 14*:745-750, 1996 and Register et al., *Plant Mol Biol 25*:951-961, 1994.

The vector further contained polynucleotides encoding signal sequences (MDWTGRFLFVVAAATGVQS; SEQ ID NO: 73) upstream from each site for insertion of the amplified immunoglobulins. The signal sequence was derived from the Kabat database (Kabat et al., Sequences of Immunological Interest, National Institutes of Health, Bethesda, Md., 1991) to allow for association of both heavy and light chain nascent polypeptides with the endomembrane system of the plant cell. In addition, the vector contained a multiple cloning site immediately downstream from the signal sequence encoding regions. The amplified regions were then ligated into the vector to encode a heavy chain polynucleotide consisting of a signal sequence, a diversity of variable region and the entire gamma constant region; and a light chain polynucleotide consisting of a signal sequence, a diversity of light chain variable region and the entire kappa constant region.

The recombinant population of vectors was then used to transform *Agrobacterium tumefasciens* as described by Ishida et al., *Nature Biotechnology 14*:745-75, 1996. The transformed *Agrobacterium* was then used to infect immature corn embryos followed by selection and regeneration of maize plants as described in Ishida et al. The resulting array of plant cells producing IgBP polypeptides (*i.e.*, IgBPs and/or IgBP components) comprised approximately 1000 plants.

To locate the position in the array of various IgBP polypeptides, samples of leaf tissue were taken from each plant. A variety of assays were performed to characterize the presence of IgBPs, the functionality of IgBPs, and the variety of IgBP components in each transformant. IgBP components (gamma chains or kappa chains) were detected by ELISA in 96 well plates. In one set of plates the capture of IgBP components was by reaction with immobilized goat anti-human kappa chain antibody and the detection was by horse radish peroxidase conjugated goat anti-human kappa chain antibody. In another set of plates the capture of IgBP components was by reaction with immobilized goat anti-human gamma chain antibody and the detection was by horse radish peroxidase conjugated goat anti-human gamma chain antibody. In a third set of plates the capture of IgBP components was by reaction with immobilized goat anti-human gamma chain antibody and the detection was by horseradish peroxidase-conjugated goat anti-human kappa chain antibody. Detection of captured IgBP components was by the procedures previously described (Antibodies: A Laboratory Manual, Harlow and Lane, eds., Cold Spring Harbor, N.Y., 1988). The results demonstrated that the majority of plants co-expressed gamma and kappa chains.

The leaf material was further evaluated for the ability of the expressed IgBPs to bind the toxin A or toxin B proteins. In these assays, ELISA plates were coated with either toxin A or B to capture the IgBPs. The ELISA procedures generally followed those previously described (Hiatt et al., *Nature 342*:76-78, 1989). Leaf extracts were incubated in the ELISA wells to allow reaction of potential immunoglobulin binding proteins with the immobilized toxins. Screening the entire array of approximately 1000 plants as described above identified 10 plants expressing functional anti-toxin A antibodies and 5 plants expressing functional anti-toxin B

antibodies. From these results, it appears that greater than 20% of the total polynucleotide population encoded anti-toxin A heavy or light chains. Similarly, greater than 14% of the total population encoded anti-toxin B heavy or light chains. The remaining 66% of the polynucleotides derived from the sorted cells may have encoded unrelated immunoglobulins.

To assess the variety of immunoglobulin binding protein components contained in those plants that expressed functional anti-toxins, PCR amplification of gamma or kappa transcripts were cloned into bacterial transformation vectors as described (Antibody Engineering, Carl A, Borrebaeck, Ed., W.H. Freeman and Company, New York, NY, 1995). Plasmids derived from each of the bacterial colonies were then subjected to automated DNA sequencing as described (DNA Sequencing: From Experimental Methods to Bioinformatics, L. Alphey, Bios Scientific Publishers, Oxford, England, 1997). The sequences derived from various plants revealed that in some cases more than one set of IgBP components was expressed in a single transformant.

To segregate the different polynucleotides encoding different IgBPs into different progeny plants, the transformants expressing multiple IgBP polypeptides were self-pollinated to produce seed sets. The resulting array of seeds was grown into individual plantlets, at which time portions of the leaf were removed to assay for the presence of functional IgBPs as described above. Since the heavy and light chains were co-expressed on the same vector, the genetic segregation potentially identifies functional anti-toxin heavy and light chains expressed from the same vector. This is useful since the parental plants expressing anti-toxins could have formed antibody from heavy and light chains polynucleotides on different vectors co-integrated into the same nucleus during *Agrobacterium* transformation. Of the 15 plants expressing anti-toxin antibodies, 14 produced a progeny array where some members of the array expressed anti-toxin antibodies indicating co-expression of heavy and light chains from the same vector. The single parental plant that produced progeny containing no anti-toxin antibodies presumably expressed anti-toxin heavy and light chain polynucleotides from different vectors, which co-integrated during transformation but which segregated in

the progeny. This observation is consistent with the known ability of *Agrobacterium* to introduce more than one vector into a nucleus during transformation.

Example 2

Isolation and Expression in Maize of Immunoglobulin Heavy Chain-Coding Genes And J Chain-Coding Gene

This Example illustrates the preparation of representative immunoglobulin heavy chain binding protein components for use in generating an array of immunoglobulin heavy chain binding proteins in plant cells.

RNA was prepared as described in Example 1. PCR amplification of the isolated RNA employed oligonucleotides complementary to conserved 3' and 5' regions of the variable region of the immunoglobulin heavy chain (IgG1) transcripts as previously described in Antibody Engineering, Carl A, Borrebaeck, Ed., W.H. Freeman and Company, New York, N.Y., 1995. The oligonucleotide primers introduced unique restriction sites at each end of the amplified polynucleotide to allow for ligation into a plant expression vector. The amplified portion of the immunoglobulin polynucleotides extended from the 5' end encoding the respective FR1 regions to the 3' end encoding the first six amino acids of the CH1 region. In addition, appropriate 5' and 3' restriction sites were introduced in the process of amplification. The amplified heavy chain polynucleotide chains therefore corresponded to the first six codons at the 5' end of the RNAs and the first six codons of the constant regions with the restriction sites introduced outside of the respective terminal codons. This immunoglobulin heavy chain binding protein array consisting of double stranded amplified cDNAs from the sorted cells was prepared for insertion into a plant expression vector, by converting the ends of the double stranded cDNA to the appropriate sticky ends for ligation into the vector (Molecular Cloning: A Laboratory Manual, Sambrook et al., eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 1989) by digestion with a restriction endonuclease according to the manufacturer's protocol (New England BioLabs). Synthetic polynucleotides encoding J chain were additionally prepared for ligation into the plant vector by PCR amplification as above. Finally, the constant region of an IgG-

A hybrid heavy chain (Ma et al., *Science 268*:716-719) was amplified by PCR to contain appropriate restriction sites for ligation into the plant vector downstream from the variable region amplified above.

As in Example 1, the vector allowed for the expression of two open reading frames using two ubiquitin promoters and nopaline synthase terminator as described by Ishida et al., *Nature Biotechnology 14*:745-750, 1996 and Register et al., *Plant Mol Biol 25*:951-961, 1994.

The vector further contained a polynucleotide encoding signal sequences (MDWTGRFLFVVAAATGVQS; SEQ ID NO: 73) upstream from the site for insertion of amplified DNAs and downstream from the first promoter. The signal sequence was derived from the Kabat database (Kabat et al., Sequences of Immunological Interest, National Institutes of Health, Bethesda, Md., 1991) to allow for association of heavy chain nascent polypeptides with the endomembrane system of the plant cell. In addition, the vector contained a multiple cloning site immediately downstream from the signal sequence codons at the first promoter site and immediately downstream from the second promoter site. The amplified heavy chain constant region was first ligated into the vector downstream from the signal sequence at the first promoter site. The remaining restriction sites located between the signal sequence and the heavy chain constant region were used for ligation of the amplified heavy chain variable regions prepared above. The amplified J chain (containing a signal sequence) was ligated downstream from the second promoter. The resulting vector encoded a heavy chain polynucleotide consisting of a signal sequence, a diversity of variable regions and the entire gamma-alpha hybrid constant region (Ma et al., *Science 268*:716-719); and a J chain polynucleotide containing a signal sequence.

The recombinant population of vectors was then used to transform *Agrobacterium tumefasciens* as described above. Regenerated leaf material was evaluated as described above with the exception that in one set of ELISA plates both the capture and detection antibody recognized human alpha constant regions. In another set of ELISA plates capture was with anti-alpha antibodies and detection was with anti-J chain antibodies. Evaluation of binding to toxin A or toxin B was performed

as described above where leaf extracts were incubated in the ELISA wells to allow reaction of potential heavy chain binding proteins with the immobilized toxins. Screening the entire array of approximately 1000 plants as described above identified 50 plants expressing functional binders of toxin A and 40 plants expressing functional binders of toxin B. From these results, it appears that the non-combinatorial approach to isolating binding proteins containing no light chain variable regions yields a higher proportion of transformation events expressing binding proteins compared to the combinatorial approach involving the random pairing of heavy and light chains. The ability to detect these binding proteins appears to be due to their assembly into a polymeric molecular structure with enhanced avidity for the toxins.

Example 3

Preparation of an Array of Functional Immunoglobulin Binding Proteins from Binding Protein Components by a Sexual Cross of Higher Plants

This Example illustrates the generation of an array of plants that express functional immunoglobulin binding proteins from plant populations that express different immunoglobulin binding protein components.

cDNAs encoding heavy and light chains derived from the B cell populations described above were cloned separately into vectors allowing for expression of heavy chains and light chains in different plant populations. The procedures described above were performed to generate plant transformants expressing arrays of heavy chains or light chains (immunoglobulin binding protein components). In order to form an array of immunoglobulin binding proteins, pollen from the light chain transformants was used to pollinate the heavy chain transformants to produce a seed set. The seeds were then grown into individual plantlets and subjected to the analysis for identification of immunoglobulin binding proteins described above. Results similar to Example 1 were obtained indicating that the random assortment of heavy and light chains by cloning them into different vectors followed by assembly upon crossing plants is comparable to cloning heavy and light chains into the same vector followed by assembly in primary transformants.

Example 4

Immunoglobulin Binding Protein Array in *Chlamydomonas reinhardtii*

This Example illustrates the preparation of an immunoglobulin binding protein array in *Chlamydomonas reinhardtii*.

A. Transformation techniques. Cells were transformed using the glass bead protocol as previously described (Kindle, *Proc. Natl. Acad. Sci. USA 87*:1228-1232, 1990). The transformation procedure described below uses the argininosuccinate lyase (ARG7, Debuchy et al., *EMBO J. 8*:2803-2809, 1989) as the selectable marker gene and transformants are selected as arginine prototrophs on medium not supplemented with arginine. Cells (arg$^-$, CW-15 strain, lacking a cell wall) were first grown under constant high light to mid-log phase in non-selective medium. About 5 x 10^7 cells were resuspended in 300 µl of medium, 150 µl 20% PEG-8000 (5% final concentration) and 0.3 g of acid-washed, heat-sterilized glass beads were then added. Ten µg of test plasmid DNA, along with 10 µg of pARG7.8 DNA (plasmid DNA containing ARG7), was added to the cells and the mixture was then agitated for 10 seconds with a vortex mixer at the highest setting. Alternatively, 5 mg of heat-sterilized silicon carbide whiskers was used and vortexed for 90 seconds (Dunahay, *BioTechniques 15*:452-460, 1993). The transformation mixture was then spread onto selective agar plates. Plates were placed under light and colonies were visible to the naked eyes after 7-10 days. All plasmid DNA used was linearized prior to use. A transformation with pARG7.8 DNA alone was always included as a control.

B. Expression of an Immunoglobulin Binding Protein Array. The IgG array isolated by PCR from the B cells selected above (anti-toxin A and B) were introduced into bacterial vectors (pBluescript). These vectors were used to transform bacteria and generate a library of polynucleotides encoding immunoglobulin binding protein components (gamma and kappa chains). The sequences of the variable regions of 100 polynucleotides (50 gamma chains and 50 kappa chains) was determined by automated sequence analysis. Variable regions of each of these polynucleotides were synthesized to reflect the codon bias of *Chlamydomonas*. The synthesis strategy

introduced unique Sac and Xho restriction sites at the ends of each DNA. Likewise, a single kappa constant region and gamma constant region (sequence derived from Kabat et al.) were synthesized with *Chlamydomonas* codon bias. The synthetic kappa and gamma constant region DNAs were first restricted and ligated into separate Xho-EcoR1 restricted pBluescript vectors to create separate vectors containing synthetic immunoglobulin constant region polynucleotides. These vectors were cloned and purified and were then used as the recipients for Sac-Xho restricted synthetic variable region DNAs to create an array of separate gamma and kappa polynucleotides encoding the diverse variable regions and constant regions of full length gamma and kappa-encoding polynucleotides. These vectors were used to transform bacteria from which plasmids were purified and automated sequence analysis was used to identify the original 100 variable regions comprising the arrays. These plasmids were mixed and the Sac-EcoR1 restriction fragments were subcloned into the pARG7 vector. The pARG7 vector contained a signal sequence derived from the Kabat database (MDWTGRFLFVVAAATGVQS; SEQ ID NO: 73) upstream from Sac-EcoR1 sites. The ligation product created full length heavy and light chain polynucleotides capable of co-expression with a signal sequence to direct polypeptide synthesis to the endomembrane system of the cell. Colonies on agar plates were tested for expression of antibodies using a nitrocellulose lift assay. Nitrocellulose circles were placed on the agar plates containing colonies on order to adsorb sufficient antibody onto the paper to allow for functional detection.

Example 5

Expression of Heavy Chain Binding Protein Genes in Insect Cells

This example illustrates the generation of an array of insect cells that express functional heavy chain binding proteins.

The heavy chain and J chain PCR products derived from B cells as described above were additionally manipulated to allow for expression in insect cells. In general, the double stranded cDNA from the selected B cells described above was prepared for insertion into an insect expression vector, by converting the ends of the

double stranded cDNA to the appropriate sticky ends for ligation into the vector (Molecular Cloning: A Laboratory Manual, Sambrook et al., eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. 1989) by digestion with a restriction endonuclease according to the manufacturer's protocol (New England Biolabs). The vector (p2Bac, Invitrogen) allowed for the expression of both heavy and J chains using the P_{P10} and the P_{PH} promoters respectively and appropriate polyadenylation sequences (BGH and PH respectively, Invitrogen). The vector therefore was potentially capable of co-expression of both a heavy chain and a J chain. The vector further contained a polynucleotide encoding a signal sequence derived from the Kabat database (Kabat et al., Sequences of Immunological Interest, National Institutes of Health, Bethesda, Md.,1991) to allow for association of both heavy and J chain nascent polypeptides with the endomembrane system of the insect cell. The signal sequence was located immediately upstream from the first polylinker which contained convenient restriction sites for introduction of additional heavy chain elements. Downstream from the first polylinker, polynucleotides encoding the constant region of the human mu heavy chain were inserted. The mu heavy chain region obtained by amplification by PCR, corresponded to codons encoding the last five amino acids and stop codon at the 3' end of the RNA, as well as the first six codons of the constant region. The heavy chain V region DNA, encoding the heavy chain portion of the combining sites recognizing toxins A or B, was ligated into the first polylinker sight after introduction of the signal and constant region elements. Polynucleotides encoding the entire J chain including signal sequence containing appropriate restricted 3' and 5' ends to allow ligation into the second polylinker site were introduced into the vector last. The vector consequently encoded a heavy chain polynucleotide consisting of a signal sequence, a variable region and an mu constant region; and a J chain polynucleotide consisting of a signal sequence and the entire coding region of a human J chain. An additional vector was constructed which encoded the constant region of the human kappa light chain downstream from a unique promoter (P_{PH} in pMelBac A, Invitrogen) and an insect signal sequence.

The resulting recombinant vectors as well as linearized AcNPV BV DNA (Pharmingen) were used to co-infect Sf insect cells using procedures

recommended by the manufacturer to generate an array of viral plaques on agar plates. Each plaque potentially contained the virions encoding unique heavy chain binding proteins as well as the unique heavy chain binding proteins themselves. Virus particles from approximately 1000 of the plaques were used to infect High Five™ insect cells (Invitrogen) in an array of 96 well plates (~10^6 cells per well). 72 hours post-infection cells and supernatant were harvested, brought to 1 mL total volume, and 10-100 µL were analyzed for composition and functionality. Identification of desired heavy chain and binding functionality can utilize any of a variety of techniques known to those of skill in the art such as ELISA, immunoblotting, Western blotting, immunoprecipitation and such, all of which are described in Harlow and Lane, "Using Antibodies: A Laboratory Manual", Cold Spring Harbor Press, Cold Spring Harbor, New York, 1999.

Analysis of the antigen binding capability of each supernatant by ELISA identified a population of heavy chain binding proteins recognizing either toxin A or toxin B. Approximately 5% of the wells containing transformed insect cells expressed detectible binding proteins recognizing toxin A and 3% expressed binding proteins recognizing toxin B. These results are similar to those of Example 2.

Example 6

Expression of Heavy Chain Binding Protein Genes in Mammalian Cells

This example illustrates the generation of an array of mammalian cells that express functional heavy chain binding proteins.

The heavy chain and J chain PCR products derived from B cells as described above were additionally manipulated to allow for expression in baby hamster kidney (BHK) cells. In general, the double stranded cDNAs from the selected B cells described in Example 1 was engineered to be expressed as human mu isotypes as described in Example 6 and was further prepared for insertion into the Sindbis expression vector, using methods describe above. The vector (pSinRep5, Invitrogen) was therefore engineered to allowed for the expression of heavy and J chains using P_{SG} promoters and appropriate polyadenylation sequences. The vector therefore was potentially capable of co-expression of both a heavy chain and a J chain.

The vector population was digested with an appropriate restriction enzyme. The resulting recombinant linearized vectors were used to generate capped RNA transcripts using the Sindbis Expression System kit. The transcripts were then used to infect BHK cells. The supernatant medium from the infected cells containing recombinant virus particles was then used identify the expressed heavy chain binding proteins. Identification of desired heavy chain can binding proteins used the techniques described above.

Analysis of the antigen binding capability of each supernatant by ELISA identified a population of heavy chain binding proteins recognizing either toxin A or toxin B. Approximately 4% of the wells containing transformed BHK cells expressed detectible binding proteins recognizing toxin A and 3% expressed binding proteins recognizing toxin B. These results are similar to those of Example 2.

From the foregoing, it will be appreciated that, although specific embodiments of the invention have been described herein for the purpose of illustration, various modifications may be made without deviating from the spirit and scope of the invention.